The Fourier Transforms of Probability Distributions

Lectures
By
AUREL WINTNER

The Fourier Transforms of Probability Distributions

Lectures
By
AUREL WINTNER
Professor of Mathematics, The Johns Hopkins University

BALTIMORE
1947

Copyright 1947
AUREL WINTNER

Printed in U. S. A.
Edwards Brothers, Inc., Ann Arbor, Michigan, 1947

These pages are based on lecture notes taken by Dr. F. W. Light during the academic year 1942 - 1943. The notes were typewritten at that time in their present form, in order to be mimeographed. Because of war conditions, this had to be postponed until now. Such advances as may have been made in the meantime are not incorporated.

11. 28. 46 A.W.

CONTENTS

Sections	Chapter Heading	Formulae	Pages
1-5	Distributions	(1)-(9)	1-11
6-10	Transforms	(10)-(29)	12-26
11-18	Moments	(30)-(47)	27-39
19-25	Convolutions	(48)-(81)	40-53
26-33	Convergent convolutions	(82)-(112)	54-72
34-41	Convergence in measure	(113)-(148)	72-92
42-48	Semi-groups	(149)-(180)	93-109
49-56	Inversions	(181)-(216)	110-126
57-68	Examples	(217)-(279)	127-156
69-75	Projections	(280)-(315)	157-180
	References		181-185

Distributions

1. Let ϕ be a field of sets E which are subsets of the line X : $-\infty < x < \infty$, and suppose that every interval is in ϕ. Clearly, ϕ contains the Borel field. It is also clear that the Borel field is a ϕ.

Let there be given on ϕ a (non-negative) measure μ such that the μ-measure of X is 1. The μ-measurability of a set can then be defined to mean that the set is contained in ϕ. This does not preclude the possibility that there exists a field more extensive than ϕ to which μ can be extended from ϕ.

It will be necessary to consider various measures, μ', μ'',... simultaneously, and so it will be convenient that the fields ϕ', ϕ'',... associated with them should be chosen to be identical. This can be accomplished by choosing ϕ to be the Borel field for every μ. This will always be done in the sequel. Thus the decision as to whether a set is or is not measurable becomes independent of the choice of the measure μ.

Accordingly, E from now on will always denote an arbitrary Borel set on the line X, and μ will be any function $\mu = \mu_E$, of E satisfying the following conditions:

(1) $\mu_E \geqq 0$; (2) $\mu_{E+F+\dots} = \mu_E + \mu_F + \dots$; (3) $\mu_X = 1$,

where E,F,... is an arbitrary sequence of mutually disjoint Borel sets and X is the set of all real numbers x.

By a distribution (on the line X : $-\infty < x < \infty$) will be meant any function $\mu = \mu_E$ of E satisfying (1), (2), (3). The name is due to the fact that the measure, μ_E, of a given E can be interpreted as the probability that a point x of X is in E.

2. Let $\lambda = \lambda_E$ be a function defined for every Borel set E in such a way that (1), (2) and the condition $\mu_X < \infty$, a condition not implying (3), are satisfied by $\mu = \lambda$.

Clearly, $\lambda_I \leqq \lambda_E \leqq \lambda_C$ for every E, if I denotes the interior, and C the closure, of E. Let an E be called a continuity set of λ if $\lambda_I = \lambda_C$. For instance, if E is an interval (closed, open or neither), and if $x = a$ and $x = b$ denote its end-points, then E is not a continuity set of λ if and only if either $\lambda_a > 0$ or $\lambda_b > 0$ (the points a,b are thought of as Borel sets). It is clear that X is a continuity set of every λ.

The last remark implies that, if $\mu^{(1)}, \mu^{(2)},\ldots$ is a sequence of distributions (i.e., if (1), (2), (3) are satisfied by $\mu = \mu^{(n)}$ for $n = 1,2,\ldots$), and if there exists a λ such that, as $n \rightarrow \infty$, the value $\mu^{(n)}_E$ tends to the value λ_E whenever E is a continuity set of λ, then $\mu = \lambda$ satisfies not only (1) and (2) but (3) as well, i.e., λ must then be a distribution.

On the other hand, if $\mu^{(1)}, \mu^{(2)}, \ldots$ is a sequence of distributions such that the limit of $\mu^{(n)}_E$ as $n \rightarrow \infty$ tends to a limit for every fixed Borel set E, then the limit, when considered as function of E, need not be a distribution. This is illustrated by the example in which the distribution $\mu^{(n)}$ is the one attaining the value $\mu^{(n)}_x = 1$ for that Borel set which consists of the single point $x = n$. The value $\mu^{(n)}_E$ then tends to 0 or to 1 according as E does not or does contain aninfinity of the points $x = 1,2,\dots$.

Conversely, a sequence of distributions $\mu^{(1)},\mu^{(2)},\dots$ can be such that there exists a distribution μ for which $\mu^{(n)}_E$ tends to μ_E for every continuity set E of μ but not for every Borel set E. An example to this effect results if $\mu^{(n)}$ is so chosen that, on the one hand, every interval is a continuity set of $\mu^{(n)}$ (i.e., if $\mu^{(n)}_x = 0$ for every point x; cf. above) and, on the other hand, $\mu^{(n)}$ attains the value $\mu^{(n)}_x = 1$ for the Borel set consisting of the interval $0 < x < 1/n$. In fact, if μ denotes the distribution for which μ_E is 0 or 1 according as the point $x = 0$ is not or is in the Borel set E, then $\mu^{(n)}_E$ obviously tends or does not tend to μ_E according as E is an arbitrary continuity set of E or consists of the Borel set represented by the point $x = 0$.

In view of the above pair of examples, the notion of a convergent sequence of distributions $\mu^{(1)},\mu^{(2)},\dots$ will have to be defined by the following requirement: There exists on X a Borel measure μ such that $\mu^{(n)}_E$ tends to μ_E whenever E is a continuity set of μ. In this, and only in this, case will the notation $\mu^{(n)} \quad \mu$ be used. Since X is always a continuity set, it follows that $\mu^{(n)} \quad \mu$ cannot hold unless μ is a distribution.

3. Let f(x) be a Baire function (of the position x) on X and let μ

be a distribution. Then, if $f(x)$ is real-valued, the set of points x satisfying $f(x) < a$, where a is any real number, is a Borel set and has therefore a μ-measure.

Hence, if $f(x)$ is any bounded Baire function on X, the Borel-Lebesgue integral $\int f(x)\,d\mu$ is defined and has, by (3), an absolute value not exceeding the least upper bound of $|f(x)|$. Here and in the sequel, any integration, $\int$, in which the range E is not marked, extends over the Borel set E consisting of the whole line, $E = X$.

If $f(x)$ is a Baire function which, instead of being bounded, is non-negative, $\int f(x)\,d\mu$ still exists but can be ∞. It is understood that "non-negative" always means "real and non-negative".

If $f_1(x), f_2(x), \ldots$ is a sequence of non-negative Baire functions on X, then, according to Fatou,

$$\int \liminf_{n\to\infty} f_n(x)\,d\mu \leqq \liminf_{n\to\infty} \int f_n(x)\,d\mu.$$

This implies that, if every $f_n(x)$ is a non-negative Baire function on X, then

$$\int f(x)\,d\mu \leqq \liminf \int f_n(x)\,d\mu \tag{4*}$$

whenever $f_n(x)$ tends to $f(x)$ almost everywhere on X, i.e., whenever $f_n(x) \to f(x)$, where $f(x) \leqq \infty$, holds for every x not contained in a set of μ-measure 0. It follows from (4*) by classical reasoning that, if the assumption $f_n(x) \geqq 0$ of (4*) is replaced by $|f_n(x)| <$ const., then (4*) can be refined to

$$\int f_n(x)\,d\mu \longrightarrow f(x)\,d\mu. \tag{4}$$

In other words, (4) holds for every uniformly bounded sequence of Baire functions $f_n(x)$ which tend to $f(x)$ almost everywhere on X. This implies that, if $f_1(x)$, $f_2(x)$,... are uniformly bounded Baire functions on an arbitrary Borel set E and tend to $f(x)$ almost everywhere on E, then

$$(4') \qquad \int_E f_n(x)\, d\mu \longrightarrow \int_E f(x)\, d\mu.$$

In fact, (4') follows from (4) by placing $f_n(x) = 0$ and $f(x) = 0$ for every x contained in the complement, $X - E$, of E.

Fatou's inequality has a dual, corresponding to the case in which the measure function, rather than the point function, varies with n. In particular, (4*) is paralleled by the inequality

$$(5^*) \qquad \int f(x)\, d\mu \leqq \liminf \int f(x)\, d\mu^{(n)},$$

if $f(x)$ is a non-negative Baire function and $\mu^{(1)}, \mu^{(2)}$,... any sequence of distributions satisfying $\mu^{(n)} \longrightarrow \mu$ in the sense defined at the end of § 2. On the other hand, it is not true that, corresponding to the assumptions of (4),

$$(5) \qquad \int f(x)\, d\mu^{(n)} \longrightarrow \int f(x)\, d\mu$$

follows from $\mu^{(n)} \longrightarrow \mu$ for every bounded Baire function $f(x)$. In fact, if $\mu^{(n)}$ and μ are the distributions defined in the second example of §2, then $\mu^{(n)} \longrightarrow \mu$ is satisfied but, if $f(x) = 0$ for every $x \neq 0$ and $f(0) = 1$, the integrals on the left and on the right of (5) have the value 0 and 1 respectively. However, (5) follows from $\mu^{(n)} \longrightarrow \mu$, if $f(x)$ is a continuous bounded function on X.

In order to prove this, it is sufficient to show that, if $\mu^{(n)} \longrightarrow \mu$, there exist arbitrarily large numbers $a > 0$ such that, if $E = E_a$ denotes the interval $-a \leqq x \leqq a$, then E is a continuity set of μ and

$$(5') \qquad \int_E f(x)\, d\mu^{(n)} \longrightarrow \int_E f(x)\, d\mu$$

holds for every $f(x)$ that is continuous on E. In fact, (5) then follows by subtraction, since on the one hand, $|f(x)| <$ const. on X and, on the other hand, $\mu^{(n)}_{X-E} \to \mu_{X-E}$ as $n \to \infty$ holds when the continuity interval, $E = E_a$, of μ is fixed, while $\mu_{X-E} = 1 - \mu_E \to 0$ holds as $a \to \infty$ in $E = E_a$. But the interval E_a is a continuity set of μ whenever both points $x = \pm a$, thought of as Borel sets, are of μ-measure 0 (cf. § 2), and (1), (2), (3) imply that the set of the points x which have a positive measure μ_x is at most enumerable. Consequently, it is sufficient to ascertain that (5') follows from $\mu^{(n)} \to \mu$ under the assumptions that E is a closed, bounded, continuity interval of μ and $f(x)$ is continuous on E. But the truth of (5') under these assumptions readily follows from the uniform continuity of $f(x)$ on E and from the definition of $\mu^{(n)} \to \mu$ at the end of § 2. (In order to see this, it is sufficient to approximate $f(x)$ by a majorizing and by a minorizing sequence of uniformly convergent step-functions, chosen in such a way that every interval of constancy of each of the step-functions is a continuity set of μ.)

It may be noted that (5') does not follow from $\mu^{(n)} \to \mu$, if E is an arbitrary closed, bounded interval on which $f(x)$ is continuous. In fact, if $f(x) = 1$ for every x, then (5) claims that $\mu^{(n)}(E)$ tends to the limit $\mu(E)$; a claim which does not follow from the assumption $\mu^{(n)} \to \mu$ (cf. § 2).

4. For a given distribution μ, let $S = S(\mu)$ denote the set of points x having the property that the μ-measure μ_J of any open interval J containing the point x is distinct from 0. It is clear from (1),(2),(3) that the set S is closed and not vacuous, and that $\mu_S = 1$.

It was pointed out in the proof of (5') that the set of points x each of which has a measure μ_x distinct from 0 is at most enumerable. Let $P = P(\mu)$ denote the set of these points. In contrast to the set S, the set P need not be closed and can be vacuous.

Let $S = S(\mu)$ be called the spectrum, and $P = P(\mu)$ the point spectrum, of μ. Clearly, P is a subset of S.

A distribution μ is called continuous if every Borel set E is a continuity set of μ (in the sense defined at the beginning of § 2). In particular, every point x must then be a continuity set, which implies that $\mu_E = 0$ for every $E = x$, the interior of every $E = x$ being the empty set (and therefore, by (1), (2), (3), a set of μ-measure 0). Accordingly, if μ is continuous, then its P is vacuous. It is easily verified from (1), (2),(3) that this condition not only is necessary but is sufficient as well in order that μ be continuous. In other words, μ is continuous if and only if every enumerable set is of μ-measure 0.

A distribution μ is called purely discontinuous in the other extreme case, namely,in case there exists an enumerable set having the same μ-measure as the whole x-line. It is clear from (3) and from the definition of P, that μ is purely discontinuous if and only if $\mu_P = 1$.

Correspondingly, μ is continuous if and only if $\mu_P = 0$. In fact, it is clear from (1), (2) and from the definition of P, that, since P is at most enumerable, μ_P is 0 if and only if P is vacuous.

While the above definitions involve only the measure μ itself, the following notions are properties of a distribution μ with reference to another measure, namely, with reference to the ordinary, i.e. "Euclidean", Borel measure (a measure which, incidentally, satisfies only the first two of the conditions (1), (2), (3) for a distribution, since it assigns to the Borel set $a < x < b$ the measure $b - a$).

A distribution μ is called purely singular if it is continuous and such that there exists a Borel set Z of Euclidean measure 0 for which $\mu_Z = \mu_X$, where X is the whole x-line. In other words, μ is called purely singular if $\mu_P = 0$ and $\mu_Z = 1$, where P denotes the point spectrum of μ and Z is a suitable set of Euclidean measure 0. Since $\mu_S = 1$ holds for every distribution, every continuous distribution having a spectrum S of Euclidean measure 0 is purely singular. On the other hand, if μ is purely singular, then S can be an interval (or, for that matter, S can be identical with X; cf.§63 below). This implies that the set Z occurring in the definition of a purely singular μ cannot in general be chosen to be a closed set, since S need not be nowhere dense. Consequently, there is no reasonable way of defining the set Z as a unique functional $Z = Z(\mu)$ of an arbitrary purely singular distribution μ. In fact, Z can be replaced by any set differing from Z by any enumerable set, since, μ being continuous, not only the Euclidean measure but also the μ-measure of every enumerable set is 0.

The other extreme case of a continuous distribution is represented by the case of absolute continuity. A distribution μ is called absolutely continuous if $\mu_E = 0$ holds for every Borel set E of Euclidean measure 0. Since P is at most enumerable, the condition $\mu_P = 0$, i.e. the continuity of μ, is a necessary (but not, of course, a sufficient) condition for the absolute continuity of μ. According to Lebesgue's form of the Fundamental Theorem of Calculus, a distribution μ is absolutely continuous if and only if there exists on X a Baire function f such that

$$(6) \qquad \mu_E = \int_E f(x)\, dx$$

holds for every Borel set E, where the integration refers to the Euclidean measure. The function $f = f_\mu(x)$ is called the density of $\mu = \mu_E$; it is uniquely determined by the distribution if and only if x-sets of Euclidean measure 0 are disregarded. In this sense, the validity of both (1) and (6) for every E implies that the density is a non-negative function of x on X. According to (3) and (6), this non-negative function has a finite integral, $\int f(x)\, dx = 1$, over the line X.

Let a distribution μ be called pure if it is either absolutely continuous or, if it is not absolutely continuous, it is either purely singular or purely discontinuous. According to Lebesgue, every distribution μ admits of a unique decomposition into pure terms,

$$(7) \qquad \mu_E = p^I \mu^I{}_E + p^{II} \mu^{II}{}_E + p^{III} \mu^{III}{}_E,$$

where E is an arbitrary Borel set, p^I, p^{II}, p^{III} are non-negative constants whose sum is unity, $\mu^I, \mu^{II}, \mu^{III}$ are distributions belonging to the three pure types respectively, and all three terms on the right of (7) are uniquely determined by the given distribution $\mu = \mu_E$.

5. If $\mu = \mu_E$ is any distribution on the line X, let $\mu(x)$, where $-\infty < x < \infty$, denote the μ-measure of the half-line $E = E^x$ consisting of those points of X which are on the left of the point x. Then it is clear from (1), (2), (3) that $\mu(x)$ is a monotone function satisfying

$$(8) \qquad \mu(-\infty) = 0, \ \mu(\infty) = 1 \text{ and } \mu(x - 0) = \mu(x)$$

(the last relation, stating that $\mu(x)$ is continuous from the left at every point x, follows from the fact that x itself was excluded from E^x). Conversely, if $\mu(x)$, $-\infty < x < \infty$, is any given monotone function satisfying (8), it is easily verified from the definition of arbitrary sets E contained in Borel's field, that there exists a unique set-function $\mu = \mu_E$ satisfying (1), (2), (3) and attaining the value $\mu(x)$ for every half-line E^x.

In view of this one-to-one correspondence, not only every function μ_E satisfying (1), (2), (3) on Borel's field, but also every monotone function $\mu(x)$ satisfying (8) on the x-line, will be called a distribution on X. Correspondingly, the Borel-Lebesgue integral with reference to the measure μ_E, as defined at the beginning of §3, is identical with the Borel-Stieltjes integral with reference to the monotone function $\mu(x)$,

$$(9) \qquad \int f(x)\, d\mu = \int_{-\infty}^{\infty} f(x)\, d\,\mu(x), \qquad (d\mu = \mu_{dE}; \ \int = \int_X),$$

whenever $f(x)$ is a Baire function for which either integral exists (e.g., for which $f(x) > -\text{const.}$).

It is clear that the value of the set-function μ_E for the Borel set E consisting of the point x is $\mu(x + 0) - \mu(x - 0)$. Hence, a point is not a continuity set of the set-function μ_E if and only if it is a discontinuity point of the point-function $\mu(x)$. In particular, μ is a continuous distribution in the sense of § 4 if and only if $\mu(x)$ is continuous on X.

For an arbitrary distribution, the function $\mu(x)$, being non-decreasing, has a finite, non-negative, derivative $\mu'(x)$ almost everywhere in Euclidean measure, and the integral $\int \mu'(x)\, dx$ over the line X has, by (8), a value not exceeding 1 (Lebesgue). In fact, $\int \mu'(x)\, dx$ is 1 if and only if the point-function $\mu(x)$ is absolutely continuous on X. This will be the case if and only if the set-function corresponding to μ is absolutely continuous in the sense of § 4, in which case (8) holds for the density $f = \mu'$ and for every Borel set E.

It is easy to prove that, if ${\mu^{(1)}}_E$, ${\mu^{(2)}}_E, \ldots$ and μ_E are distributions on X, then $\mu^{(n)} \longrightarrow \mu$ holds in the sense defined at the end of § 2 if and only if the corresponding point-functions, $\mu^{(n)}(x)$ and $\mu(x)$, have the property that $\mu^{(n)}(x)$ tends to the limit $\mu(x)$ at every continuity point x of μ.

Transforms

6. Let y be a point of a line $-\infty < y < \infty$. Then $f(x) = e^{ixy}$, being a bounded Baire function on the line $-\infty < x < \infty$, is a function for which the integral (9) exists. This integral, a function of y on the line $-\infty < y < \infty$, will be called the transform of the distribution which is represented by μ on the line $-\infty < x < \infty$. Thus

$$(10) \qquad \mu^{o}(y) = \int e^{ixy}\, d\mu(x); \quad -\infty < y < \infty, \qquad \left(\int = \int_{-\infty}^{\infty}\right),$$

if $\mu^{o}(y)$ denotes the transform of $\mu(x)$. It is clear from (1) and (3) that

$$(11) \qquad |\mu^{o}(y)| \leqq 1; \qquad (12) \quad \mu^{o}(0) = 1,$$

and, from (12), that $\int (1 - e^{ixy})\, d\mu(x) = 1 - \mu^{o}(y)$, hence

$$(13) \qquad \int (1 - \cos yx)\, d\mu(x) = \text{real part of } 1 - \mu^{o}(y).$$

It is also seen from (10) that

$$(14) \qquad \mu^{o}(y)\, \mu^{o}(-y) = |\mu^{o}(y)|^{2}.$$

Let ε denote the distribution whose spectrum S consists of the single point $x = 0$, i.e., for which

$$(15) \qquad \varepsilon(x) = 0 \text{ if } x < 0 \text{ and } \varepsilon(x) = 1 \text{ if } x > 0; \text{ so that } \varepsilon^{o}(y) \equiv 1,$$

by (10). The distribution $\varepsilon(x)$ will be called the unit distribution.

If $\mu(x)$ is any distribution, and if c is a real number, let ${}^{c}\mu(x)$ denote the distribution, $\mu(x - c)$, which results if the origin of the x-axis is translated by c. Then it is seen from (10) that ${}^{c}\mu^{o}(y) = e^{icy}\mu^{o}(y)$. In particular, from (15),

$$(16) \qquad {}^{c}\varepsilon^{o}(y) = e^{icy}; \text{ so that } {}^{c}\mu^{o}(y) = {}^{c}\varepsilon^{o}(y)\, \mu^{o}(y).$$

In contrast to the definition of the distribution ${}^{c}\mu$, let the distribution ${}_{c}\mu$ be defined by changing the units along the x-axis, and, if $c < 0$, by a subsequent transposition of the x-axis, as follows: For a given distribution $\mu(x)$, let ${}_c\mu(x)$ denote the distribution $\mu(x/c)$ or the unit distribution $\varepsilon(x)$ according as $c > 0$ or $c = 0$; while if $c < 0$, let ${}_c\mu(x)$ denote the distribution which is equal to $1 - \mu(x/c)$ when x is a continuity point of ${}_c\mu$ and, in accordance with (8), is continuous from the left at every x. It is seen from (10) that the transform of the distribution ${}_c\mu$ is given by

$$(17) \qquad {}_c\mu^{\circ}(y) = \mu^{\circ}(cy)$$

in both cases $c > 0$, $c < 0$; and (15), when compared with (12), shows that (17) holds in the limiting case $c = 0$ also.

It is clear from the definition of the two linear sheaves, ${}^c\mu$ and ${}_c\mu$, just defined for every distribution μ, that the distribution ${}^a_b\mu$ is not identical with the distribution ${}^b_a\mu$ (not even in the case $a = b$).

From the case $c = -1$ of (17) and from (14),

$$(18) \qquad {}_{-1}\mu^{\circ}(y) = \mu^{\circ}(-y) = \text{complex-conjugate of } \mu^{\circ}(y).$$

According to (16), the absolute value of the transform of every translation, ${}^c\varepsilon$, of the unit distribution is constant. It turns out that this is a characteristic property of the distributions ${}^c\varepsilon$. More than this will be proved by showing that a distribution μ must be of the form ${}^c\varepsilon$, if $\mu^{\circ}(y) = \mu^{\circ}(0)$ holds for a sequence of y-values which are not contained in an equidistant sequence, $y = \alpha + \beta k$, $(k = 0, \pm 1, \pm 2, \ldots)$. Hence, it is more than sufficient to show that, if there exists a single non-vanishing y, say $y = y_0$, satisfying $\mu^{\circ}(y_0) = \mu^{\circ}(0)$, then there exists a number $x_0 = x_0(y_0)$ such that the spectrum of $\mu(x)$ is contained in the periodic sequence $x = x_0 + 2\pi k/y_0$. According to (12) and (11), this means that the absolute value of a transform cannot attain its maximum at two points unless the distribution has a spectrum contained in a periodic sequence.

In order to prove this, suppose that corresponding to μ there exists a $y_0 \neq 0$ satisfying $|\mu^o(y_0)| = 1$. Then (16) implies that, if μ is replaced by a suitably translated distribution, ${}^c\mu$, it can be assumed that $\mu^o(y_0) = 1$. But then (13) shows that, the function $1 - \cos(xy_0)$ of x being continuous and non-negative, $1 - \cos(xy_0) = 1$ must hold for every x contributing to the integral $\int(1 - \cos xy_0)d\mu$; cf. (1). This means that μ must be purely discontinuous, with a point spectrum contained in the sequence formed by the roots x of the equation $1 = \cos(xy_0)$. This completes the proof of all the statements made above.

7. It will now be shown that for the transformation (10) of a distribution μ into μ^o the following uniqueness theorem holds:

$$\mu_1{}^o(y) \equiv \mu_2{}^o(y) \text{ only when } \mu_1(x) \equiv \mu_2(x). \tag{19}$$

If (10) and (8) are applied to both monotone functions μ_1, μ_2, it is seen, by placing $\beta = \mu_1 - \mu_2$, that (19) is equivalent to the following statement:

(19*) if $\int |d\beta(x)| < \infty$ and $\int e^{ixy} \, d\beta(x) \equiv 0$, then $\beta(x - 0) = \text{const.}$

The proof of (19*) proceeds as follows: By assumption, $\beta(x)$, $-\infty < x < \infty$, is a function of bounded variation satisfying $\int f d\beta = 0$ for every function $f(x) = e^{ixy}$, where y is an arbitrary real constant. Hence, $\int f d\beta = 0$ holds for any trigonometric polynomial $f(x) = ae^{ixy} + be^{ixz} + \dots$, and so, since $\int |d\beta| < \infty$, for any function f(x) which is a uniform limit of such polynomials. It follows therefore from Weierstrass' approximation theorem, that $\int f d\beta = 0$ whenever f(x) is a continuous periodic function. It will now be shown that $\int f d\beta = 0$ holds for $f = f_J$ also, if $f_J(x)$ denotes the function which is 1 or 0 according as x is or is not in J, where J is any bounded interval on the x-axis. This will complete the proof of (19*), since it is clear that $\int f_J d\beta = 0$ cannot hold for every J unless the function $\beta(x)$, which is of bounded variation, is constant on the set of its continuity points.

Let J be fixed and let p be any positive number so large that the interval $-p < x \leqq p$ contains J. Let $f_J^p(x)$ denote the function which has the period 2p and attains, if $-p < x \leqq p$, the value 1 or 0 according as x is or is not in J. Clearly, f_J^p can be represented as the limit of a uniformly bounded sequence of continuous functions each of which has the period 2p. Since $\int f d\beta = 0$ holds for each of these approximating functions, it follows from $\int |d\beta| < \infty$ that $\int f_J^p d\beta = 0$. But

$f_J^{\,p}(x) \longrightarrow f_J(x)$ as $p \longrightarrow \infty$ holds for every x. It follows therefore from $\int |d\beta| < \infty$ and $|f_J^{\,p}| \leqq 1$ that $\int f_J d\beta = 0$. This completes the proof of (19*).

This proof of (19) needs the assumption of (19) for $-\infty < y < \infty$. Actually, there does not exist any Const. $< \infty$ having the property that $\mu_1 = \mu_2$ is implied by the truth of $\mu_1^{\,0}(y) \equiv \mu_2^{\,0}(y)$ for $|y| <$ Const. In view of (17), it is sufficient to prove this negation for the case Const. = 1, a case in which the truth of the negation follows by considering, for instance, the even functions g(y) and h(y) which are defined for $-\infty < y < \infty$ as follows: The function g(y) equals $1 - |y|$ for $-1 < y < 1$ and has the period 2, while h(y) is g(y) or 0 according as $-1 < y < 1$ is or is not satisfied. Thus $h(y) = \mathrm{Max}(0, 1 - |y|)$ for every y, and so a Fourier inversion (or an integration of Dirichlet's factor of discontinuity) shows that $h(y) = \int f(x) e^{ixy} dx$, where $-\infty < x < \infty$, is satisfied for $-\infty < y < \infty$ by an elementary continuous function f(x) which is nowhere negative. Since $\int f(x)\, dx = 1$ in view of $h(0) = 1$, it follows that (6) defines an absolutely continuous distribution. If the latter is called μ_1, a comparison of (10) with the definition of f(x) shows that $\mu_1^{\,0}(y) = h(y)$ for $-\infty < y < \infty$. Similarly, if the Fourier constants of the function $g(y) = g(y + 2)$ are evaluated, it is found that every a_k is non-negative in the Fourier series $g(y) = \sum a_k e^{i\pi ky}$, where $k = 0, \pm 1, \ldots$. Thus there exists a purely discontinuous distribution μ by means of which g(y) is representable in the form (10) for $-\infty < y < \infty$. If this μ is called μ_2, then $\mu_1^{\,0}(y) = \mu_2^{\,0}(y)$ for $-1 < y < 1$, since $h(y) = g(y)$ for $-1 < y < 1$. But the distributions μ_1, μ_2 are not identical, since μ_1 is continuous, while μ_2 is not.

8. In view of this example, it is of interest that there exist certain particular distributions μ_1 having the property that, if any distribution μ_2 satisfies the relation $\mu_1{}^{o}(y) = \mu_2{}^{o}(y)$ for every y of sufficiently small absolute value, then μ_2 satisfies this relation for every y, and so $\mu_2 = \mu_1$.

Such a distribution μ_1 is, for instance, the unit distribution, ε. This is implied by the results found at the end of § 6. Since this property of ε will become of fundamental importance, it is worthy of a more direct proof (a proof which, incidentally, appears to be the oldest of its type).

According to (15), the assertion to be proved is that, if $\mu^{o}(y) = 1$ holds in a vicinity of the point $y = 0$, then $\mu = \varepsilon$. Hence, it is more than sufficient to prove that

$$\text{(20)} \qquad \text{if } \mu^{o}(y) = 1 + o(y^2) \text{ as } y \longrightarrow 0, \text{ then } \mu = \varepsilon.$$

But (13) and the assumption of (20) imply that $y^{-2}\int(1 - \cos yx)d\mu \longrightarrow 0$ as $y \longrightarrow 0$. It follows therefore from $1 - \cos yx \geqq 0$ that, as $y \longrightarrow +0$,

$$y^{-2}\int_{-1/y}^{1/y}(1 - \cos yx)\,d\mu(x) = o(1); \text{ hence } \int_{-1/y}^{1/y} x^2\,d\mu(x) = o(1),$$

since $\alpha^{-2}(1 - \cos \alpha) > \text{const.} > 0$ if $0 < \pm\alpha < 1$. The second relation of the last formula line, when compared with (1) and (15), completes the proof of $\mu = \varepsilon$.

Two similar consequences of (13), to be used in the sequel, can be formulated as follows: If μ is any distribution, then

$$y^2 \int_{-r}^{r} x^2 d\mu(x) < 3\,|1 - \mu^{o}(y)| \text{ for } |y| < 1/r, \tag{21}$$

where $r > 0$ is arbitrary, and, again for every $r > 0$,

$$\left(\int_{-\infty}^{\infty} - \int_{-r}^{r}\right) d\mu(x) < 2r \int_{-1/r}^{1/r} |1 - \mu^{o}(y)|\, dy. \tag{22}$$

First, since the real part of a complex number z has an absolute value not greater than $|z|$, it is clear from (13) that

$$\int_{-r}^{r} (1 - \cos yx)\, d\mu(x) + \left(\int_{-\infty}^{\infty} - \int_{-r}^{r}\right) (1 - \cos yx)\, d\mu(x) \leqq |1 - \mu^{o}(y)|, \tag{23}$$

where y and r (> 0) are arbitrary. Both terms on the left of (23) are non-negative. If the second term on the left is omitted, (21) follows from (23) by observing that $\alpha^2 \leqq 3(1 - \cos \alpha)$ for $-1 \leqq \alpha \leqq 1$. If the first term on the left is omitted, (22) follows from (23) by integration over the interval $-1/r < y < 1/r$, since

$$\int_{-1/r}^{1/r} (1 - \cos xy)dy = \frac{2}{r}\left(1 - \frac{r}{x}\sin \frac{x}{r}\right) \text{ and } 2(1 - \sin 1) > \tfrac{1}{2}.$$

An obvious adaption of the proof of (20) shows that

$$\text{if } \mu^{o}(y) = 1 + O(y^2) \text{ as } y \to 0, \text{ then } \int x^2 d\mu(x) < \infty. \tag{24}$$

In fact, it is clear that, if the o-assumption of (20) is replaced by the O-assumption of (24), then the o(1) in the formula line following (20) can be replaced by O(1); which proves that $\int x^2 d\mu(x) < \infty$.

It should be mentioned that (24) becomes false if y^2, x^2 are replaced by $|y|$, $|x|$ respectively. In other words, $\int |x| \, d\mu(x) = \infty$ is compatible with $\mu^0(y) = 1 + O(|y|)$ as $y \rightarrow 0$. In fact, not even $\mu^0(y) = 1 + o(|y|)$ as $y \longrightarrow 0$ ensures that $\int |x| \, d\mu(x) < \infty$.

In order to prove this, let $\mu(x)$ be the absolutely continuous distribution for which the density, $\mu'(x)$, is 0 or const./$(x^2 \log x)$ according as $|x| < 2$ or $|x| > 2$ (the numerical value of const. > 0 is assigned by the condition $\int \mu'(x)dx = \int d\mu(x) = 1$). Clearly, $\int |x| d\mu(x) = \infty$ in this case. On the other hand, from (12) and (10),

$$0 < \frac{1 - \mu^0(y)}{\text{const.}} = \int_2^\infty \frac{1 - \cos yx}{x^2 \log x} \, dx.$$

It follows that $\mu^0(y)$ is an even function and that, since $0 \leqq 1 - \cos \alpha \leqq \mathrm{Min}(1, \alpha^2)$ for $-\infty < \alpha < \infty$, the difference $1 - \mu^0(y)$ is majorized by a constant multiple of

$$\int_2^{1/y} \frac{y^2 x^2}{x^2 \log x} dx + \int_{1/y}^\infty \frac{1}{x^2 \log x} \, dx = y^2 O \ \frac{1}{y} /\log \frac{1}{y} \ + O \ \ y/\log \frac{1}{y}$$

$= O(-y/\log y) = o(y)$, as $y \longrightarrow +0$.

9. The uniqueness theorem, (19), states that the knowledge of the transform, μ^o, of a distribution, μ, is sufficient for the determination of μ itself. There arises the question as to the existence of a corresponding completeness theorem; that is, of a theorem which assumes appropriate convergence properties of the transforms of a sequence of distributions, and asserts that, if the transforms have these properties, the distributions themselves must converge (in the sense defined at the end of § 2). It will now be shown that such a completeness theorem can be formulated as follows: If the transforms $\mu_1^o, \mu_2^o, \ldots$ of a sequence $\mu_1, \mu_2, \ldots$ of distributions have the property that there exists on the y=line a function $\phi = \phi(y)$ which is continuous at the point $y = 0$ and is such that $\mu_n^o(y) \rightarrow \phi(y)$ as $n \longrightarrow \infty$ holds at every point y, then there exists a distribution μ satisfying $\mu_n \rightarrow \mu$ as $n \longrightarrow \infty$; moreover, the transform of this μ is precisely the given limit function:

$$\mu^o(y) = \phi(y). \tag{25}$$

The proof of this completeness theorem will require of the function $\phi(y)$ only the property that

$$r \int_{-1/r}^{1/r} |1 - \phi(y)|\, dy \rightarrow 0 \text{ as } r \rightarrow \infty. \tag{26}$$

Since $\mu_n^o(y) \rightarrow \phi(y)$ for every y, it is seen from the case $\mu = \mu_n$ of (10), (11) and (12), that $\phi(y)$ is a bounded, L-measurable function satisfying $\phi(0) = 1$, and so it is clear that the continuity of $\phi(y)$ at $y = 0$ is sufficient for (26). Actually, it will turn out in §10 that the milder assumption, namely (26), implies the continuity of $\phi(y)$ at $y = 0$ and, for that matter, at every y.

However, some restriction of the behavior of $\phi(y)$ at $y = 0$ is certainly needed in the completeness theorem. In other words, if it is only assumed that there exists a function $\phi(y)$ satisfying $\mu_n^{\circ}(y) \longrightarrow \phi(y)$ for every y, then there need not exist a distribution μ satisfying $\mu_n \longrightarrow \mu$. An example to this effect results by placing $\mu_n(x) = \mu_1(x/n)$ and choosing $\mu_1 = \mu_1(x)$ to be the absolutely continuous distribution for which the density $\mu_1'(x)$ is 1 or 0 according as x is or is not in the interval $0 < x < 1$. Then it is clear that the limit of $\mu_n(x)$ as $n \longrightarrow \infty$ exists and is 0 for every x; so that, since the constant 0 is not a distribution, there does not exist a distribution μ satisfying $\mu_n \longrightarrow \mu$. However, $\mu_1^{\circ}(y) = i(1 - e^{iy})/y$, by (10), and $\mu_n^{\circ}(y) = \mu_1^{\circ}(ny)$, by (17); so that, if $\phi(y)$ denotes the function which is 1 or 0 according as $y = 0$ or $y \neq 0$, then $\mu_n^{\circ}(y) \longrightarrow \phi(y)$ holds for every y. Correspondingly, (26) is violated.

In order to prove the completeness theorem formulated above, use will be made of Helly's selection theorem, according to which every uniformly bounded sequence of non-decreasing functions contains a subsequence which converges to a non-decreasing function at every continuity point of the latter. Thus, if $\mu_1(x), \mu_2(x), \ldots$ is any sequence of distributions, it contains a subsequence, say $\mu_{(1)}(x), \mu_{(2)}(x), \ldots$, such that $\mu_{(n)}(x)$ tends, as $n \longrightarrow \infty$, to $\mu(x)$ for every x that is not a discontinuity point of μ, where μ is a certain non-decreasing function. Clearly, μ can be normalized so that it becomes continuous from the left at every x. However, the preceding example implies that μ will not in general be a distribution.

But it will now be shown that μ must be a distribution whenever the assumptions of the completeness theorem are satisfied. Since $\mu_1, \mu_2, \ldots$ are distributions, it will of course be sufficient to show that

$$(27^*) \quad \limsup_{n \to \infty} \left(\int_{-\infty}^{\infty} - \int_{-r}^{r} \right) d\mu_n(x) \to 0 \text{ as } r \to \infty .$$

It follows therefore from (22) that it will be sufficient to show that

$$(27) \quad \limsup_{n \to \infty} r \int_{-1/r}^{1/r} | 1 - \mu_n{}^{\circ}(y) | dy \to 0 \text{ as } r \to \infty .$$

To this end, choose a positive number ε and determine, corresponding to it, a positive $q = q_\varepsilon$ satisfying

$$r \int_{-1/r}^{1/r} | 1 - \phi(y) | \, dy < \varepsilon \text{ for } r > q_\varepsilon .$$

The existence of such a q_ε is the assumption (26). The other assumption is that $\mu_n{}^{\circ}(y) \to \phi(y)$ as $n \to \infty$ holds for every y. According to Lebesgue's theorem on term-by-term integration, this implies that, if $r > 0$ is arbitrarily fixed,

$$\lim_{n \to \infty} r \int_{-1/r}^{1/r} | 1 - \mu_n{}^{\circ}(y) | dy = r \int_{-1/r}^{1/r} | 1 - \phi(y) | dy,$$

since, in view of the case $\mu = \mu_n$ of (10) and (11), the functions $\mu_n{}^{\circ}(y)$ are L-integrable and uniformly bounded. Clearly, the last two formula lines entail the truth of (27).

This proves that the sequence $\mu_1, \mu_2, \ldots$ contains a subsequence $\mu_{(1)}$, $\mu_{(2)}, \ldots$ for which there exists a distribution μ satisfying $\mu_{(n)} \to \mu$. This implies, of course, that every subsequence of $\mu_1, \mu_2, \ldots$ contains a subsequence which converges to a distribution. It will now be shown that, if a subsequence of $\mu_1,\mu_2,\ldots$ tends to a distribution, then the latter is independent of the choice of the convergent subsequence. According to the principle of the excluded middle, this will prove that the sequence $\mu_1,\mu_2,\ldots$ itself tends to a distribution.

In order to prove that μ is independent of the subsequence whose limit it is, let $\mu_{(1)},\mu_{(2)},\ldots$ be any subsequence of $\mu_1,\mu_2,\ldots$ satisfying $\mu_{(n)}\longrightarrow\mu$ for a certain distribution μ. Then, if (5) is applied to the sequence $\mu_{(1)}, \mu_{(2)},\ldots$ and to the bounded continuous function $f(x) = e^{ixy}$, where y is arbitrarily fixed, it follows from the definition (10), that $\mu_{(n)}{}^{o}(y)\longrightarrow\mu^{o}(y)$ holds for every y. On the other hand, $\mu_{(n)}{}^{o}(y)\longrightarrow\phi(y)$ for every y, since, by assumption, $\mu_n{}^{o}(y)\longrightarrow\phi(y)$ for every y. Consequently, μ satisfies (25) for every y. But the function $\phi(y)$ is independent of the choice of the subsequence $\mu_{(1)},\mu_{(2)},\ldots$ of $\mu_1,\mu_2,\ldots$, since $\mu_n{}^{o}(y)\longrightarrow\phi(y)$. It follows therefore from (25) and from the uniqueness theorem (§ 7), that μ is independent of the choice of the subsequence $\mu_{(1)},\mu_{(2)},\ldots$ for which $\mu_{(n)}\longrightarrow\mu$.

This concludes the proof of the completeness theorem, formulated before (25).

10. It is easy to see that, if μ and $\mu_1,\mu_2,\ldots$ are arbitrary distributions,

(28) $\mu_n \longrightarrow \mu$ if and only if $\overset{o}{\mu_n}(y) \longrightarrow \mu^o(y)$ for every y.

This is a much easier statement than the completeness theorem, since, in contrast to the latter, the assertion of (28) does not contain an existence statement. Correspondingly, (28) can be proved without having recourse to (22).

If $\mu_n \longrightarrow \mu$, then, if (5) is applied to the bounded, continuous function $f(x) = e^{ixy}$, where y is fixed, it is seen that one of the assertions of (28) is clear from the definition (10). In view of the completeness theorem of § 9, the remaining assertion of (28) will be proved if it is shown that the function (25) is continuous at $y = 0$. In other words, it is sufficient to show that the transform $\mu^o(y)$ of every distribution μ is continuous at $y = 0$.

Actually, the transform $\overset{o}{\mu}(y)$ of every μ is continuous at every y; moreover, every $\mu^o(y)$ is uniformly continuous for $-\infty < y < \infty$.

In order to prove this, let $\mu^o(y)_r$ denote, for every positive number r, the contribution of the interval $-r < x < r$ to the transform (10); so that

$$(29) \qquad \mu^o(y)_r = \int_{-r}^{r} e^{ixy}\, d\mu(x).$$

Since the contribution of the interval $-r \leqq x \leqq r$ to $\int d\mu(x) = 1$ tends to 1 as $r \longrightarrow \infty$, and since $| e^{ixy} | = 1$, it is clear that $\mu^o(y)_r \longrightarrow \overset{o}{\mu}(y)$ as $r \longrightarrow \infty$ holds uniformly for $-\infty < y < \infty$, and that it is sufficient to prove the uniform continuity of the function (29) for $-\infty < y < \infty$, if r is fixed. But, if y and h are arbitrary real numbers, (29) shows that

$$|\mu^o(y+h)_r - \mu^o(y)_r| = \left|\int_{-r}^{r} e^{ixy}(e^{ixh}-1)\,d\mu(x)\right| \leqq \int_{-r}^{r} |e^{ixh}-1|\,d\mu(x).$$

Since the last integral is independent of y and tends, as $h \to 0$, to 0 for every fixed r, the proof is complete.

It may be mentioned that, if $\mu_n \to \mu$, the transform $\mu_n{}^o(y)$ must tend to the transform $\mu^o(y)$ uniformly on every fixed bounded y-interval. In fact, if $\mu_n \to \mu$, it is clear from the definition of $\mu_n \to \mu$ (cf. the end of § 2), that (27*) is satisfied. Hence it is clear from the definitions (29), (10), that it is sufficient to prove this: If $\mu_n \to \mu$, then there exist certain arbitrarily large values of r such that $\mu_n{}^o(y)_r \to \mu^o(y)_r$ as $n \to \infty$ holds uniformly on every fixed bounded y-interval.

It is easy to see that there exist arbitrarily large values of r possessing this property. In fact, a positive number r will have this property whenever both $x = r$ and $x = -r$ are continuity points of the monotone function μ, that is, whenever the interval $-r < x < r$ is a continuity set of $\mu = \mu_E$ (cf. § 2). Then (5'), where $\mu^{(n)} = \mu_n$ in the present notations, is applicable to the set E represented by the interval $-r < x < r$ and to the function $f(x) = e^{iyx}$, where y is arbitrary. This means, by (29), that the relation $\mu_n{}^o(y)_r \to \mu^o(y)_r$ holds for every fixed y. In order to prove that it holds uniformly on every fixed bounded y-interval, it is sufficient to ascertain that the functions $\mu_1{}^o(y)_r$, $\mu_2{}^o(y)_r, \ldots$, where r is fixed, are equicontinuous (Arzelà) on every such y-interval; for instance, that the functions $\mu_n{}^o(y)_r$ of y have uniformly bounded derivatives. But (29) shows that these derivatives exist and have the absolute values

$$\left| \int_{-r}^{r} ixe^{ixy}\, d\mu_n(x) \right| \leqq \int_{-r}^{r} |x|\, d\mu_n(x) \rightarrow \int_{-r}^{r} |x|\, d\mu(x),$$

by (5'). Since the last integral is independent of y and n, the proof is complete.

The above results, when combined with the completeness theorem (§ 9), can be formulated as the following continuity theorem of the transform (10): For a given sequence of distributions $\mu_1, \mu_2, \ldots$ on the x-line, there exists a distribution μ satisfying $\mu_n \longrightarrow \mu$ if and only if the sequence of the transforms $\mu_1^{\circ}(y)$, $\mu_2^{\circ}(y), \ldots$ is uniformly convergent on every fixed bounded y-interval; in which case the function $\phi(y) = \lim \mu_n^{\circ}(y)$ is precisely the transform $\mu^{\circ}(y)$ of the distribution $\mu = \lim \mu_n$.

Moments

11. For a non-negative integer k and for a distribution μ, the k-th moment of μ is defined to be the real number

(30) $$(\mu)_k = \int x^k d\mu(x), \text{ if } \int |x|^k \, d\mu(x) < \infty .$$

If $\int |x|^k \, d\mu = \infty$, then $(\mu)_k$ is considered as undefined. Thus it is clear from (30) that the existence of the (k + 1)-th moment of μ is sufficient for the existence of the k-th. It is also clear that the 0-th moment exists for every μ, since (3) means that $(\mu)_0 = 1$.

Let μ and k be fixed. If $\int |x|^k \, d\mu < \infty$ and $m \leqq k$, then the integral obtained by m-fold formal differentiation of (10) obviously is uniformly convergent for $-\infty < y < \infty$. Hence, $D_y^m \mu^o(y)$, where $D_y^m = d^m/dy^m$, exists and is represented by

(31) $$D_y^m \mu^o(y) = i^m \int e^{ixy} x^m d\mu(x), \text{ if } \int |x|^m \, d\mu < \infty .$$

In particular, from (30),

(32) $$D_y^m \mu^o(0) = i^m (\mu)_k, \text{ if } \int |x|^m \, d\mu < \infty .$$

It is also clear from (31) that the k-th derivative of the transform is continuous for every y. In particular, it is continuous at y = 0. It follows therefore from Taylor's formula that

(33) $$\mu^o(y) = \sum_{m=0}^{k} i^m (\mu)_m \, y^m/m! + o(|y|^k) \text{ as } y \longrightarrow 0,$$

since the assumption of (32) is satisfied for every index $m \leqq k$, if it is satisfied for m = k.

Accordingly, if μ has a k-th moment, then the transform of μ behaves at y = 0 in such a way that there exist k + 1 constants c_m satisfying

(33') $$\mu^o(y) = \sum_{m=0}^{k} c_m y^m + o(|y|^k) \text{ as } y \longrightarrow 0.$$

The converse of this result does not hold. In fact, if μ is the distribution considered at the end of § 8, then $\mu^{o}(y) = 1 + o(|y|)$ but $\int |x| d\mu = \infty$; so that (33') is satisfied for $k = 1$ (by $c_0 = 1$, $c_1 = 0$), although $(\mu)_k = (\mu)_1$ does not exist.

12. The index k is odd in the counter-example just given. It turns out that the converse of the result of § 11 is true whenever the index k is even. In fact, not only (33'), but also the weaker assumption which replaces the o-remainder by the corresponding O-remainder, implies the truth of the converse, if k is of the form 2n. In other words, the existence of $(\mu)_{2n}$ is not only sufficient (§ 11), but is necessary as well, for the existence of constants c_m satisfying

$$(34) \qquad \mu^{o}(y) = \sum_{m=0}^{2n-1} c_m y^m + O(y^{2n}) \text{ as } y \longrightarrow 0.$$

In fact, it is useless to specify in (34) the term $c_m y^m$ of degree $m = k$, where $k = 2n$, since this specification is canceled by the $O(y^{2n})$ in (34). Thus the assertion to be proved contains, among other things, the curious statement that the assumption of 2n constants satisfying (34) necessitates the existence of a (2n + 1)-st constant satisfying (33') for k = 2n.

If $n = 0$, the assertion to be proved is trivial, since $(\mu)_0 = 1$ for every μ. Hence, it is sufficient to show that, if the assertion is true for every index less than n, then it is true for n also. To this end, suppose that a μ satisfies (34) for a given n. Then (34) is of course satisfied if n is replaced by $n - 1$. Since the assertion is supposed to be true for the index n, it follows that $(\mu)_{2n-2}$ exists.

According to § 11, this implies that (33) holds for $k = 2n - 2$. It follows therefore from a comparison with (34), that $c_m = i^m(\mu)_m/m!$ if $m < 2n - 1$, where $(\mu)_m = \int x^m d\mu$, by (30). Consequently, (34) can be written in the form

$$\mu^o(y) = \sum_{m=0}^{2n-2} \int (ixy)^m/m!\ d\mu + c_{2n-1}y^{2n-1} + o(y^{2n}).$$

Hence, if the real parts are taken on both sides, (10) shows that

$$\int \cos yx\ d\mu = \sum_{k=0}^{n-1} \int (-1)^k(xy)^{2k}/(2k)!\ d\mu + ay^{2n-1} + o(y^{2n}),$$

where $a = a_{2n-1}$ denotes the real part of c_{2n-1}. Thus $\int g_n(xy)\ d\mu = ay^{2n-1} + o(y^{2n})$ as $y \rightarrow 0$, if $g_n(\alpha)$ is an abbreviation for

$$g_n(\alpha) = \cos \alpha - \sum_{k=0}^{n-1} (-1)^k \alpha^{2k}/(2k)! \tag{35}$$

According to a standard property of the partial sums of the infinite series $\cos \alpha = \sum (-1)^k \alpha^{2k}/(2k)!$, the function (35) does not change its sign for $-\infty < \alpha < \infty$. Hence, the function $\int g_n(xy)\ d\mu$ of y cannot change its sign. On the other hand, this function of y was just shown to be of the form $ay^{2n-1} + O(y^{2n})$, as $y \longrightarrow 0$, and must therefore change its sign unless $a = 0$. Consequently, $\int g_n(xy)\, d\mu = O(y^{2n})$, and so, since $g_n(\alpha)$ does not change its sign, $\int |g_n(xy)|\, d\mu = O(y^{2n})$. Since this integral extends over the whole x-line, it follows that

$$y^{-2n} \int_{-b/y}^{b/y} |g_n(xy)|\, d\mu(x) = O(1) \text{ as } y \longrightarrow +0,$$

where b is arbitrary. But it is clear from (35) that there exists for every sufficiently small $b = b(n) > 0$ a positive constant satisfying $g_n(\alpha) > \text{const.}\, \alpha^{2n}$ for $-b < \alpha < b$. Consequently, as $y \longrightarrow +0$,

$$y^{-2n} \int_{-b/y}^{b/y} \text{const.}\, (xy)^{2n}\, d\mu(x) = O(1), \text{ i.e., } \int_{-b/y}^{b/y} x^{2n}\, d\mu(x) = O(1).$$

Since $b/y \longrightarrow \infty$ as $y \longrightarrow +0$, the last relation completes the proof for the existence of $(\mu)_{2n}$.

13. It is clear that

$$\int |x|\, d\mu(x) < \infty \text{ implies } \int |x + c|\, d\mu(x) < \infty \tag{36}$$

for every real number c. This means that, if ${}^{c}\mu(x)$ denotes, as before (16), the translated distribution $\mu(x - c)$, then the existence of the first moment of $\mu = {}^{0}\mu$ implies that the first moment of ${}^{c}\mu$ exists for every c and that

$$(37) \qquad ({}^{c}\mu)_1 = (\mu)_1 + c.$$

In fact, (37) follows from (30), since $(\mu)_0 = 1$.

If $(\mu)_1$ exists, let $\mu^{\bullet}$ denote the distribution ${}^{c}\mu$ obtained from μ by the translation $c = (\mu)_1$; so that

$$(38) \qquad \mu^{\bullet}(x) = \mu(x - (\mu)_1), \text{ i.e., } (\mu^{\bullet})_1 = 0,$$

by (37). A distribution will be called "centered" if it has a vanishing first moment, i.e., if it is a distribution $\mu^{\bullet}$. A distribution μ can be made centered only if it has a first moment, since otherwise the distribution $\mu^{\bullet}$ is undefined.

14. The rule (37) states that, if $(\mu)_1$ exists, any translation of the x-axis subjects the value of $(\mu)_1$ to the same translation. This rule would not have been true if, in contrast to the agreement made after (30), the first moment $(\mu)_1$ would have been considered as defined whenever $\int x \, d\mu(x)$ exists as a principal value, i.e., whenever the contribution of the symmetric interval $-r \leqq x \leqq r$ to $\int x \, d\mu\,(x)$ tends to a limit as $r \to \infty$.

In fact, the violation of the identity (37) in case of such principal integrals is illustrated by the absolutely continuous distribution having the even density $\mu'(x) = \text{const.}(1 + x^2)^{-1}$. Correspondingly, this μ, though symmetric, is not centered, since $(\mu)_1$ does not exist.

15. Suppose that μ has a second moment $(\mu)_2 = \int x^2 d\mu < \infty$. Then obviously $\int (x + c)^2 \, d\mu < \infty$ for every c. This means that ${}^c\mu(x) = \mu(x - c)$ has a second moment, and therefore certainly a first moment. Thus, from (30),

$$({}^c\mu)_2 = (\mu)_2 + 2c(\mu)_1 + c^2, \quad , \tag{39}$$

since $(\mu)_0 = 1$. In particular, from (38) and (39),

$$(\mu^\bullet)_2 = (\mu)_2 - (\mu)_1^{\,2} = \int (x - (\mu)_1)^2 \, d\mu(x). \tag{40}$$

Clearly, the non-negative integral (40) is 0 if and only if $x = (\mu)_1$ is the only point of increase of the non-decreasing function $\mu(x)$, i.e., if and only if μ is a translation ${}^c\varepsilon$ of the unit distribution ε. It follows therefore from (40) that $(\mu^\bullet)_2 > 0$ or $(\mu^\bullet)_2 = 0$ according as μ is or is not a distribution ${}^c\varepsilon$, i.e., according as $\mu^\bullet$ is or is not the distribution ε.

In view of the remarks made with regard to (37), the proviso of the definition (30) cannot reasonably be disposed of, if k is odd. But if k is even, it will from now on be convenient to cancel the agreement made after (30), i.e. to use the notation $(\mu)_{2n} = \int x^{2n} \, d\mu$ whether $(\mu)_{2n} < \infty$ or $(\mu)_{2n} = \infty$.

If this understanding is applied to the case $2n = 2$, it is possible to define the "standard deviation" of an arbitrary distribution μ, as follows: The standard deviation, to be denoted by $[\mu]$, is non-negative and its square (the "variance" of μ) is given by

$$(41)\qquad [\mu]^2 = \begin{cases} \infty \text{ if } (\mu)_2 = \infty, \\ (\mu)_2 - (\mu)_1^2 = (\mu^\bullet)_2 \text{ if } (\mu)_2 < \infty; \end{cases}$$

cf. (40). In particular, $[\mu] = 0$ if and only if $\mu = {}^c\varepsilon$, i.e., $\mu^\bullet = \varepsilon$.

16. Let μ be a distribution possessing finite moments $(\mu)_k$ of every index k or, what is the same thing, of every even index k. According to § 12, this will be the case if and only if (33) holds for every even k, and therefore for every k; so that (32) and (31) are valid for every k and for every real y. Accordingly, all moments of a distribution $\mu(x)$ exist if and only if its transform $\mu^o(y)$ has a Taylor series

$$(42)\qquad \sum_{k=0}^{\infty} D_y^k \mu^o(0)\, y^k/k! \equiv \sum_{k=0}^{\infty} i^k (\mu)_k y^k/k!, \text{ cf. (32)},$$

where it is understood that this power series can diverge for every $y \neq 0$.

It will now be shown that, if the Taylor series (42) happens to converge for some $y \neq 0$, and therefore on some interval, say on $-q < y < q$, then the sum of (42) must be the function $\mu^o(y)$ for every y on this interval. In other words, the behavior exhibited by the Taylor series, $0 + 0y + 0y^2 + \ldots$, of the function $f(y) = \exp(-1/y^2)$ cannot occur for functions which are transforms $f(y) = \mu^o(y)$ of distributions $\mu(x)$.

In order to prove this, let $f(y)$ be a function possessing derivatives of arbitrarily high order on an interval $-q < y < q$, and let p be any positive number less than q. Then it is known that the Taylor series $\sum D_y^k f(0)y^k/k!$ converges to $f(y)$ on the interval $-q < y < q$ if and only if, for every fixed p less than q, the maximum of the absolute value of $D_y^k f(y)$ on the interval $-p \leqq y \leqq p$ is $O(k!/p^k)$ as $k \longrightarrow \infty$. In the present case, where $f(y) = \mu^0(y)$, this condition can, by (31), be written in the form

$$\text{Max}\left|\int e^{ixy}x^k d\mu(x)\right| = O(k!/p^k) \text{ as } k \longrightarrow \infty, \text{ where } |y| \leqq p < q. \tag{43}$$

Clearly, the expression on the left of (43) cannot exceed $\int |x|^k d\mu$ for any k, and, according to (32), is identical with $\int x^k d\mu$ for every even k. On the other hand, it is easily verified from Schwarz's inequality and from Stirling's formula, that the estimate

$$\int |x|^k d\mu = O(k!/p^k) \text{ as } k \longrightarrow \infty, \qquad (p < q), \tag{44}$$

must hold for every p, if it holds for every p when k is restricted to be even. But (30) shows that (44) can then be written in the form

$$(\mu)_{2n} = O((2n)!/p^{2n}) \text{ as } n \longrightarrow \infty, \qquad (p < q). \tag{45}$$

Consequently, either all or none of the three conditions (43), (44), (45) are satisfied for every p.

In particular, (43) is equivalent to (45). Hence, the series (42) converges to the sum $\mu^0(y)$ on the interval $-q < y < q$ if and only if (45) is satisfied for every p. Consequently, all that remains to be ascertained is that the mere convergence of (42) on the interval $-q < y < q$ necessitates the estimate (45) for every p. But this is obvious, since the convergence of the series (42) on the interval $-q < y < q$ implies that the terms of the series form a bounded sequence of numbers for every y contained in the interval, and therefore for every p.

17. Suppose that (42) converges for some $y \neq 0$, say for $-q < y < q$. Then the result of § 16 can be formulated as follows: Let w be a complex variable, and let y and v denote its real and imaginary parts respectively. Then there exists in the circle $|w| < q$ a regular function, say $\mu^{0}(w)$, which becomes identical with the transform (10) on the real diameter of the circle.

It will now be shown that the function $\mu^{0}(w)$ thus defined in the circle $|w| < q$ always admits of an analytic continuation which is regular in the whole strip $|v| < q$ about the y-axis, and that, in addition, $\mu^{0}(w)$ is a bounded function on every narrower strip, $|v| < p$.

First, it is seen from $(\mu)_{2n} = \int x^{2n}\, d\mu$ that, if the distribution μ and the positive number p are arbitrary, then

$$(46)\qquad \sum_{k=0}^{\infty}\int |x|^{k}\, d\mu(x)\; p^{k}/k! = \int \exp\ (p|x|)\,d\mu(x),$$

where it is understood that either side can be ∞ . Suppose that μ is such as to make the integral (46) finite for a certain p. Then it is clear that the integral which results by writing $w = y + iv$ in place of y in (10) is uniformly convergent and bounded on the strip $|v| < p$. Hence, it is sufficient to ascertain that the expression (46) is finite for every $p < q$, if (42) converges for $-q < y < q$. But it was shown in § 16 that (42) converges for $-q < y < q$ if and only if (44) holds for every $p < q$. Since this clearly is the case if and only if the series (46) converges for every $p < q$, the proof is complete.

It follows, in particular, that a transform $\mu^{o}(y)$ must be regular-analytic on the whole line $-\infty < y < \infty$ whenever it is regular-analytic on some interval $-q < y < q$. This obviously implies that, if two transforms, say $\mu_1{}^{o}(y)$ and $\mu_2{}^{o}(y)$, are identical on an interval $-q < y < q$ on which $\mu_1{}^{o}(y)$ is regular-analytic, then they must be identical for $-\infty < y < \infty$; which means, by (19), that $\mu_1 = \mu_2$. This criterion contains a substantial generalization of the fact deduced in § 8 from (20), namely of the fact that $\mu = \varepsilon$ if $\mu^{o}(y)$ is constant on some interval $-q < y < q$.

18. The above results have a certain, somewhat loose, connection with Stieltjes' problem of determinateness, and lead to the following sufficient condition for uniqueness:

$$\text{(47)} \qquad 2n\text{-th root of } (\mu)_{2n} < \text{const. } n.$$

(A somewhat sharper sufficient condition, namely

$$\text{(47')} \qquad \sum_{n=0}^{\infty} 2n\text{-th root of } (\mu)_{2n}{}^{-1} = \infty,$$

is supplied by the theory of quasi-analytic functions).

Stieltjes found that there exist sequences of numbers which are moment sequences of several distributions. In other words, there exist pairs μ_1, μ_2 of distributions for which $(\mu_1)_k$ and $(\mu_2)_k$ exist for every k and $(\mu_1)_k = (\mu_2)_k$

holds for every k, although $\mu_1 \neq \mu_2$. Let a distribution μ be called of the determined type if $(\mu)_k$ exists for every k and if $\mu = \lambda$ follows for every distribution λ for which $(\lambda)_k$ exists and equals $(\mu)_k$ for every k. A necessary and sufficient condition in order that a given μ be of the determined type results from the spectral theory of the infinite matrices of the continued fractions associated with the Stieltjes transform $\int (x - z)^{-1}\, d\mu(x)$. However, this criterion is all too implicit to be "applicable" in general.

In order to prove that (47) is sufficient for a μ of the determined type, suppose that $(\mu_1)_k$ and $(\mu_2)_k$ exist and are equal for every k, and let $(\mu)_k$ denote their common value. It is clear from Stirling's formula that (47) is equivalent to the existence of a positive number p (or q) satisfying (45). According to § 16, this will be the case if and only if there exists a q such that the two power series (42) belonging to $\mu = \mu_1$ and $\mu = \mu_2$ converge for $-q < y < q$ to $\mu_1{}^{0}(y)$ and to $\mu_2{}^{0}(y)$ respectively. But these two power series are identical, since $(\mu_1)_k = (\mu_2)_k$ for every k. Hence, $\mu_1 = \mu_2$ follows from the criterion pointed out at the end of § 17.

If the regular-analyticity of $\mu^{0}(y)$ near $y = 0$ is replaced by its quasi-analyticity near $y = 0$, the generalization (47') of (47) follows by repeating, word by word, the above proof of the criterion (47).

Incidentally, no explicit examples of distributions μ appear to be given in the literature which would prove that (47), (47') and a necessary and sufficient criterion for a distribution of the determined type are actually distinct.

Convolutions

19. For any pair $\mu_1 = \mu_1(x)$, $\mu_2 = \mu_2(x)$ of distributions, let $\mu_1 * \mu_2 = \mu_1 * \mu_2(x)$ denote denote the function

$$(48) \qquad \mu_1 * \mu_2(x) = \iint\limits_{v+u<x} d\mu_1(v)\, d\mu_2(u)$$

(so that $\mu_1 * \mu_2(x) = \mu_2 * \mu_1(x)$, since (48) is symmetric). Since

$$\int\limits_{v<t} d\mu_1(v) = \int_{-\infty}^{t-0} d\mu_1(v) = \mu_1(t-0) - \mu_1(-\infty) = \mu_1(t-0),$$

it follows, by placing $t = x - u$ and $\int = \int_{-\infty}^{\infty}$, that

$$(48') \qquad \int \Big\{ \int\limits_{v<x-u} d\mu_1(v) \Big\} d\mu_2(u) = \int \mu_1(x-u-0)\, d\mu_2(u).$$

But the repeated integral on the left of (48') is identical with the double integral (48). Moreover, $\mu_1(x - u - 0)$ on the right of (48') can be replaced by $\mu_1(x - u)$; cf. (8). Consequently,

$$(49) \qquad \mu_1 * \mu_2(x) = \int \mu_1(x-u)\, d\mu_2(u).$$

Since $\mu = \mu_1$ and $\mu = \mu_2$ are non-decreasing functions satisfying (8), it is easily seen from (48) that $\mu = \mu_1 * \mu_2$ is a non-decreasing function satisfying (8). Thus (49) defines a distribution $\mu_1 * \mu_2$ for every pair of distributions μ_1, μ_2.

The distribution $\mu_1 * \mu_2$ is called the convolution of μ_1 and μ_2. It is characterized by the property that, if $f(x)$ is any Baire function which is either bounded or non-negative, then

$$(50) \qquad \iint f(v+u)\, d\mu_1(v)\, d\mu_2(u) = \int f(x)\, d(\mu_1 * \mu_2)(x)$$

(the double integral on the left extends over the whole (u,v)-plane). In fact, (50) follows from (48) by applying Fubini's theorem onthe product space of the lines $-\infty < v < \infty$, $-\infty < u < \infty$ which carry the respective measures μ_1, μ_2. Actually, Fubini's theorem was used even in the transition from (48) to (48'). Correspondingly, (50) becomes identical with (48) if $f(v + u)$ is chosen to be 1 or 0 according as $v + u$ is or is not less than x.

Since $\int d\mu_2(u) = 1$, the integral (49) can be thought of as multiplied by $1/\int d\mu_2(u)$. Then (49) means that the distribution $\mu_1 * \mu_2$ is the average, with reference to the weight function μ_2, of all translated distributions ${}^{u}\mu_1(x) = \mu_1(x - u)$ belonging to a given distribution $\mu_1(x)$.

This interpretation of the convolution of μ_1 and μ_2 seems to be unsymmetric but is not; in fact, as pointed out after (48),

$$(51) \qquad \mu_1 * \mu_2 = \mu_2 * \mu_1.$$

In addition, if μ_1, μ_2, μ_3 are three distributions, then $\mu_1 * \mu_2 * \mu_3$ is a unique distribution, since

$$(52) \qquad (\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3).$$

In fact, if $v + u$ and $\int\int d\mu_1(v)\, d\mu_2(u)$ in (48) are replaced by $v + u + t$ and $\int\int\int d\mu_1(v)\, d\mu_2(u)\, d\mu_3(t)$ respectively, the associative law (51) follows for reasons of symmetry. Another proof of (51) and (52) is contained in (53) below.

20. If (50) is applied to the bounded continuous function $f(x) = e^{ixy}$, where y is fixed, (10) shows that (50) becomes

(53) $$\mu_1^{\circ}(y)\,\mu_2^{\circ}(y) = (\mu_1 * \mu_2)^{\circ}(y).$$

Thus the product of the transforms of two distributions is again a transform, namely that of the convolution.

For instance, (14) and (18) imply that

(54) $$|\mu^{\circ}(y)|^2 = ({}_{-1}\mu * \mu)^{\circ}(y)$$

for every μ. Similarly, from (16),

(55) $${}^{c}\mu^{\circ} = ({}^{c}\varepsilon * \mu)^{\circ}, \text{ i.e. } {}^{c}\mu = {}^{c}\varepsilon * \mu,$$

by (19). This means that the operation of translating $\mu = \mu(x)$ into ${}^{c}\mu(x) = \mu(x - c)$ is equivalent to the operation of convoluting $\mu = \mu(x)$ with the translated united distribution ${}^{c}\varepsilon(x) = \varepsilon(x - c)$; a fact which is clear from (49) also.

If μ_n and λ_n, where $n = 1,2,\ldots$, are two convergent sequences of distributions, then

(56) $$\mu_n * \lambda_n \rightarrow \mu * \lambda, \text{ if } \mu_n \rightarrow \mu \text{ and } \lambda_n \rightarrow \lambda.$$

In fact, (56) follows by applying (53) to both $\mu_1 * \mu_2 = \mu_n * \lambda_n$ and $\mu_1 * \mu_2 = \mu * \lambda$, if the resulting identities are combined with the facts which result when (28) is applied to all three sequences μ_n, λ_n, $\mu_n * \lambda_n$.

More recondite than the "multiplication rule" (56) is the "division rule", according to which

(57) $$\lambda_n \rightarrow \varepsilon, \text{ if } \mu_n \rightarrow \mu \text{ and } \mu_n * \lambda_n \rightarrow \mu,$$

where ε denotes the unit distribution.

First, $(\mu_n * \lambda_n)^{\circ}(y) = \mu_n^{\circ}(y)\ \lambda_n^{\circ}(y)$, by (53). It follows therefore from (28) that, in virtue of the assumptions of (57), both limit relations $\mu_n^{\circ}(y) \rightarrow \mu^{\circ}(y)$, $\mu_n^{\circ}(y)\ \lambda_n^{\circ}(y) \rightarrow \mu^{\circ}(y)$ hold for every y. Consequently, $\lambda_n^{\circ}(y) \rightarrow 1$ holds for every y satisfying $\mu_n^{\circ}(y) \neq 0$. But $\mu_n^{\circ}(y) \neq 0$ is certainly satisfied for every y contained in a sufficiently small interval $-q < y < q$, since, on the one hand, it was shown after (29) that every $\mu^{\circ}(y)$ is a continuous function and, on the other hand, $\mu^{\circ}(0) \neq 0$, by (12). Consequently, $\lambda_n^{\circ}(y) \rightarrow 1$ for $-q < y < q$. It will now be shown that $\lambda_n^{\circ}(y) \rightarrow 1$ cannot hold for $-q < y < q$ unless it holds for $-\infty < y < \infty$. If this is shown, then the assertion of (57) follows from (28), since $1 = \varepsilon^{\circ}(y)$, by (15).

21. The proof will be such as to apply not only to the distribution ε but to any distribution, say λ, which has the property that every transform identical with the transform $\lambda^{\circ}(y)$ on an interval $-q < y < q$ is identical with $\lambda^{\circ}(y)$ for $-\infty < y < \infty$. It was shown at the beginning of § 8 that ε is such a λ, and it turned out at the end of § 17 that such a λ results as soon as $\lambda^{\circ}(y)$ is assumed to be regular-analytic at $y = 0$.

In order to facilitate references to preceding sections, let the notations λ_n, λ be changed to μ_n, μ respectively. Then the assertion is that, if μ_n and μ are distributions satisfying $\mu_n^{\circ}(y) \rightarrow \mu^{\circ}(y)$ for $-q < y < q$, and if μ has the particular property that a transform cannot be identical with $\mu^o(y)$ for $-q < y < q$ unless it is identical with $\mu^o(y)$ for $-\infty < y < \infty$ then $\mu_n{}^o(y) \longrightarrow \mu^o(y)$ must hold for $-\infty < y < \infty$ (which means, by (28), that $\mu_n \longrightarrow \mu$).

The proof consists of an adaptation of the proof used in § 9. In order to see this, apply Helly's selection theorem in the same way as before (27*). In other words, choose from the sequence of the distributions $\mu_1, \mu_2, \ldots$ a subsequence $\mu_{(1)}, \mu_{(2)}, \ldots$ which converges to a non-decreasing bounded function at every continuity point of the latter, and normalize this limit function so as to become continuous from the left at every x. Let the limit function thus normalized be denoted by $\lim \mu_{(n)}$. For the present, it is undecided whether or not the total variation of $\lim \mu_{(n)}$ is 1, i.e., whether or not $\lim \mu_{(n)}$ is a distribution.

In § 9, the proof that $\lim \mu_{(n)}$ is a distribution was based on (27), and the assumption was that $\mu_n{}^o(y) \longrightarrow \mu^o(y)$ for $-\infty < y < \infty$. However, since $1/r \longrightarrow 0$ in (27), this part of the proof remains valid under the present assumption, according to which $\mu_n{}^o(y) \longrightarrow \mu^o(y)$, instead of being valid for $-\infty < y < \infty$, is known to hold only on some interval $-q < y < q$. In other words, it follows in the same way as in § 9, that $\lim \mu_{(n)}$ is a distribution.

This means that $\mu_{(n)} \longrightarrow \lim \mu_{(n)}$ holds in the sense defined at the end of § 2. Hence, if $f(x)$, $\mu^{(n)}$, μ in (5) are replaced by e^{ixy}, $\mu_{(n)}$, $\lim \mu_{(n)}$ respectively, it follows from (10) that $\mu_{(n)}{}^o(y) \longrightarrow (\lim \mu_{(n)})^o(y)$ for $-\infty < y < \infty$. On the other hand, if $-q < y < q$, then, by assumption, $\mu_n{}^o(y) \longrightarrow \mu^o(y)$, and therefore $\mu_{(n)}{}^o(y) \longrightarrow \mu^o(y)$, the sequence $\mu_{(1)}, \mu_{(2)}, \ldots$

being a subsequence of $\mu_1, \mu_2, \ldots$. Consequently, $\lim \mu_{(n)}{}^{o}(y) = \mu^{o}(y)$ holds for $-q < y < q$. But this implies, by the particular property assumed for $\mu^{o}(y)$, that $\lim \mu_n{}^{o} = \mu$.

Accordingly, the relation $\mu_{(n)} \longrightarrow \lim \mu_{(n)}$ can be written in the form $\mu_{(n)} \longrightarrow \mu$. Since μ is given with the sequence $\mu_1, \mu_2, \ldots$ and is therefore independent of the choice of the subsequence $\mu_{(1)}, \mu_{(2)}, \ldots$, it follows, as in § 9, from the principle of the excluded middle, that the proof of $\mu_n \longrightarrow \mu$ is complete.

22. From here on to (62) below, let $f(x)$, where $-\infty < x < \infty$, denote a Baire function which is non-negative, even, and does not decrease as $|x|$ increases, i.e., which for every x and for every $r > 0$ satisfies the following conditions:

$$(58) \qquad f(x) = |f(x)| = f(|x|) \leqq f(|x| + r).$$

There will be needed two sufficient criteria, to be expressed in terms of additional restrictions on $f(x)$. The first of these criteria will ensure that

$$(59) \qquad \int f(x)\, d\mu(x) < \infty$$

holds for both $\mu = \mu_1$ and $\mu = \mu_2$ whenever it holds for $\mu = \mu_1 * \mu_2$, while the second will ensure the converse. It turns out that

$$(60) \qquad f(x + u) = O(f(x)) \text{ as } x \longrightarrow \infty \text{ for every fixed } u$$

and

(61) $f(x + u) = O(f(x) + f(u))$ as $x + u \longrightarrow \infty$

are sufficient criteria of the first and of the second kind respectively.

In other words, (58) and (60) ensure that (59) holds for both $\mu = \mu_1$ and $\mu = \mu_2$ whenever (59) holds for $\mu = \mu_1 * \mu_2$, while (58) and (61) ensure that (59) holds for $\mu = \mu_1 * \mu_2$ whenever (59) holds for both $\mu = \mu_1$ and $\mu = \mu_2$. The proofs will depend on the fact that, $f(x)$ being a non-negative Baire function, (50) is true whether (59) does or does not hold for $\mu = \mu_1 * \mu_2$ (in fact, (50) resulted by a mere application of Fubini's theorem, and is therefore valid even if either side of (50) becomes ∞ for a non-negative f).

In order to prove the first of the criteria stated, suppose that f satisfies (58) and (60), and that (59) holds for a given $\mu = \mu_1 * \mu_2$. The latter assumption means that the integral on the right of (50) is finite. Hence, if the double integral on the left of (50) is thought of as a repeated integral, it follows from $f \geqq 0$ that the interior integral on the left of (50) must be finite for at least one u. In other words, there must exist a real number c satisfying $\int f(x + c)\, d\mu_1(x) < \infty$. But then the identities (58) and the inequality (58), when combined with the estimate (60), imply that (59) holds for $\mu = \mu_1$. In view of (51), this completes the proof of the first criterion.

In order to prove the second criterion, suppose that f satisfies (58) and (61). Then the integral on the left of (50) cannot exceed a constant multiple $\int\int (f(v) + f(u))d\mu_1(v)\, d\mu_2(u)$. But, since $\int d\mu_1(v) = 1$ and $\int d\mu_2(u) = 1$, this double integral is identical with the sum of the two simple integrals represented by the cases $\mu = \mu_1$ and $\mu = \mu_2$ of the integral on the left of (59). Hence, if (59) holds for both $\mu = \mu_1$ and $\mu = \mu_2$, then the integral on the left of (50) is finite. Since the integral on the right of (50) is identical with the case $\mu = \mu_1 * \mu_2$ of the integral on the left of (59), the proof of the second criterion is complete.

23. As an application, let $f(x) = |x|^k$, where $k > 0$. Then (58) and (60) are satisfied. In addition, (61) is satisfied, since $|x + u|^k < |2x|^k + |2u|^k$. Consequently, both criteria are applicable. Accordingly, $\int |x|^k \, d\mu(x) < \infty$ holds for $\mu = \mu_1 * \mu_2$ if and only if it holds for both $\mu = \mu_1$ and $\mu = \mu_2$, where k is any positive constant. If k is an integer, it follows that $\mu_1 * \mu_2$ has a finite k-th moment if and only if both μ_1 and μ_2 possess finite k-th moments; in which case application of (32) to all three distributions gives

$$(62) \qquad (\mu_1 * \mu_2)_k = \sum_{m=0}^{k} \binom{k}{m} (\mu_1)_m (\mu_2)_{k-m}; \quad (\mu)_0 = 1,$$

if the product (53) is differentiated k times.

For instance, $(\mu_1 * \mu_2)_1$ exists if and only if both $(\mu_1)_1$ and $(\mu_2)_1$ exist; in which case (62) gives

$$(63) \qquad (\mu_1 * \mu_2)_1 = (\mu_1)_1 + (\mu_2)_1$$

and implies therefore, by (38), that

$$(64) \qquad (\mu_1 * \mu_2)^{\cdot} = (\mu_1)^{\cdot} * (\mu_2)^{\cdot}$$

Similarly, $(\mu_1 * \mu_2)_2$ is finite if and only if both $(\mu_1)_2$ and $(\mu_2)_2$ are finite; in which case (62) shows that

(65) $$(\mu_1 * \mu_2)_2 = (\mu_1)_2 + 2(\mu_1)_\mu(\mu_2)_1 + (\mu_2)_2,$$

and so, in particular

(66) $$(\mu_1^{\cdot} * \mu_2^{\cdot})_2 = (\mu_1^{\cdot})_2 + (\mu_2^{\cdot})_2$$

in virtue of (64) and (68). In the contrary case, the first line of the definition (41) gives

(67) $$[\mu_1 * \mu_2]^2 = [\mu_1]^2 + [\mu_2]^2,$$

since then both sides of (67) are ∞. It follows therefore from (66) and from the second line of the definition (41), that the standard deviations satisfy (67) for any pair of distributions μ_1, μ_2.

It is clear from the proof given in § 22 for the sufficiency of condition (61) that (61) can be replaced by

(6. bis) $$f(x + u) = O(f(x)\ f(u)) \text{ as } x + u \longrightarrow \infty.$$

Obviously, this condition and both (58) and (60) are satisfied if $f(x) = \exp px$, where p is a positive constant. Consequently, the integral (46) is finite for $\mu = \mu_1 * \mu_2$ if and only if it is finite for both $\mu = \mu_1$ and $\mu = \mu_2$. But it was shown in § 17 that the integral (46) is finite for every $p < q$ if and only if the transform $\mu^{o}(y)$ is regular-analytic for $-q < y < q$; in which case the transform admits of a regular analytic continuation $\mu^{o}(y + iv)$ into the strip $|v| < q$. Accordingly, this will be the case for the transform of $\mu_1 * \mu_2$ and for a given q if and only if it is the case for the transforms of both μ_1 and μ_2 and for the same q. The "only if" part of this fact is by no means obvious from (53).

24. If A and B are two (not necessarily disjoint) sets on the x-axis, let their "vectorial sum", A (+) B, be defined as the set of those numbers x which are representable in at least one manner as $x = a + b$, where a and b are restricted to be in A and in B, respectively. It is understood that

(68) A (+) B is vacuous if either A or B is.

If both A and B are single points, a and b, then the vectorial addition, a (+) b, is equivalent to the ordinary addition, $a + b$. If B is a single point, b, then A (+) b is the set obtained from the set A by the translation b. Generally,

(69) $$A \ (+) \ B = B \ (+) \ A.$$

If C is a third set, the set A (+) B (+) C is unique, i.e.,

(70) $$(A \ (+) \ B) \ (+) \ C = A \ (+) \ (B \ (+) \ C).$$

If both sets A,B are closed, the set A (+) B need not be closed. In fact, if A and B consist of the points $a = n$ and $b = n^{-1} - n$ respectively, where $n = 1,2,\ldots,$ then A and B are closed but A (+) B contains the points n^{-1} without containing the point $0 = \lim n^{-1}$. However, it is readily shown that, if A and B are closed, then A (+) B must be closed as soon as either A or B is bounded.

It is easy to see that, if $S = S(\mu)$ denotes, as at the beginning of § 4, the spectrum of a distribution μ, then

(71) $$S(\mu_1 * \mu_2) = \text{closure of } S(\mu_1) \ (+) \ S(\mu_2),$$

and that the corresponding rule for point spectra is

$$(72) \qquad P(\mu_1 * \mu_2) = P(\mu_1)\ (+)\ P(\mu_2).$$

In fact, (71) and (72) follow directly from (48) and from the definitions of $S(\mu)$ and $P(\mu)$ in §4, if the set-functions $\mu = \mu_E$, used in §4, are replaced by the corresponding point-functions $\mu = \mu(x)$, defined in §5. (The discrepancy between (71) and (72) is due, of course, to the fact that $S(\mu)$ always is, while $P(\mu)$ need not be, a closed set; cf. §4).

In addition,

$$(73) \qquad \lim S(\mu_n) \text{ contains } S(\mu), \text{ if } \mu_n \longrightarrow \mu.$$

Here $\lim A_n$ can be defined, for a given sequence of sets $A_1, A_2, \ldots$, to be the set of those points x for which $a_n \longrightarrow x$ as $n \longrightarrow \infty$ holds for suitable points $a_1 = a_1(x)$, $a_2 = a_2(x), \ldots$ of $A_1, A_2, \ldots$ respectively. Thus $\lim A_n$ always exists (but can be vacuous even if none of the A_n is). With this definition of $\lim A_n$, the truth of (73) is clear from the definitions of the symbols $S(\mu)$ and $\mu_n \longrightarrow \mu$, introduced at the beginning of §4 and at the end of §2 respectively.

Trivial examples show that the converse of (73) is false, i.e., that $\mu_n \longrightarrow \mu$ does not imply the relation $\lim S(\mu_n) = S(\mu)$; and that nothing can be said if S is replaced by P (in fact, neither of the sets $\lim P(\mu_n)$, $P(\lim \mu_n)$, where $\mu_n \longrightarrow \mu$, need be vacuous when the other set is vacuous).

25. For a given distribution μ on the line $-\infty < x < \infty$ and for every positive number s, let $d = d(s)$ denote the least upper bound of $|\mu(x') - \mu(x'')|$ when $|x' - x''| < s$. Thus $d(s)$ is a non-decreasing function satisfying $d(\infty) = 1$ and $d(+0) \geqq 0$, where the sign of equality, i.e. $d \longrightarrow 0$ as $s \longrightarrow 0$, holds if and only if the distribution μ is continuous (in fact, since $\mu(-\infty) = 0$ and $\mu(\infty) = 1$, it is clear that $\mu(x)$ cannot be continuous for every x unless it is uniformly continuous for $-\infty < x < \infty$). Incidentally, it is easily seen that, whether μ is continuous or not, $d(s + t)$ cannot exceed $d(s) + d(t)$, where $s > 0$ and $t > 0$ are arbitrary.

Since $|x' - x''| = |(x' - u) - (x'' - u)|$ for every u, it is clear that $|\mu(x' - u) - \mu(x'' - u)| < d(s)$ whenever $|x' - x''| < s$. Hence it is seen from (49) that, if $\mu_1 = \mu$, $\mu_2 = \lambda$, then the difference $|\mu * \lambda(x') - \mu * \lambda(x'')| \leqq \int |\mu(x' - u) - \mu(x'' - u)|\, d\lambda(u)$, where $\int d\lambda(u) = 1$, cannot exceed $d(s)$. In other words, if $d_{12}(s)$. $d_1(s)$, $d_2(s)$ denote the functions $d(s)$ belonging to $\mu = \mu_1 * \mu_2$, μ_1, μ_2 respectively, then, for every $s > 0$,

$$(74) \qquad d_{12}(s) \leqq d_1(s), \text{ and so } d_{12}(s) < d_2(s),$$

by (51). This implies that $d_{12}(+0)$ cannot exceed the smaller of the numbers $d_1(+0)$, $d_2(+0)$. This means that the greatest saltus of $\mu_1 * \mu_2$ cannot exceed the greatest saltus of μ_1 or that of μ_2, where it is understood that the greatest saltus of $\mu(x)$ is the maximum of $\mu(x + 0) - u(x - 0)$ for $-\infty < x < \infty$ and can, therefore, be 0. For the latter case, (72) supplies the following refinement of what thus follows from (74):

(75) The distribution $\mu_1 * \mu_2$ is continuous if and only if at least one of the two distributions μ_1, μ_2 is.

In fact, (75) follows from (72) and (68), since, according to § 9, the point spectrum $P(\mu)$ is vacuous if and only if the distribution μ is continuous.

The inequalities (74) represent just one manifestation of a fairly general, though somewhat vague, principle, according to which the function $\mu_1 * \mu_2$ is at least as "smooth" as the "smoother" of the two functions μ_1, μ_2. Other illustrations of this principle will be seen below. Still another manifestation of it can be formulated as follows:

(76) If either μ_1 or μ_2 is absolutely continuous, then $\mu_1 * \mu_2$ is absolutely continuous.

In view of (75), it is of interest that the converse of (76) is false; in fact, it is so fundamentally false that

(77) $\mu_1 * \mu_2$ can be absolutely continuous when both μ_1 and μ_2 are purely singular (cf. § 4).

An example proving (77) will be given in § 63 below.

The proof of (76) can be based on the formula line preceding (74). However, the following proof of (76) is preferable, since it supplies, besides the assertion of (76), the explicit form of the density of $\mu_1 * \mu_2$, if either μ_1 or μ_2 has a density (cf. § 4).

If $\mu_1(x)$ is any distribution, then, since it is a monotone function, it has a finite, non-negative derivative $\mu_1'(x)$ almost everywhere (here and in the sequel, the "almost everywhere" refers to the Euclidean measure on the x-axis). If $\mu_2(x)$ is another distribution, then, since $\mu_1'(x) \geqq 0$

almost everywhere, the Lebesgue-Stieltjes integral

$$(78) \qquad \phi(x) = \int \mu_1'(x - u)\, d\mu_2(u)$$

exists, and is non-negative, for every x, where it is understood that (78) can be ∞ for some x (possibly not only almost everywhere). In any case, an application of Fubini's theorem to (78) gives

$$(79) \qquad \int_{-\infty}^{x} \phi(x)\, dv = \int \left\{ \int_{-\infty}^{x} \mu_1'(v - u)\, dv \right\} d\mu_2(u).$$

This holds for any pair of distributions μ_1, μ_2. Suppose now that, while μ_2 is left arbitrary, μ_1 is absolutely continuous. Then the interior integral on the right of (79) is $\mu_1(x - u)$ for every x and every u. Hence, if (79) is compared with (49), it follows that

$$(80) \qquad \mu_1 * \mu_2(x) = \int_{-\infty}^{x} \phi(v)\, dv$$

for every x. But the function on the left of (80), and therefore integral on the right of (80), if finite for every x. Hence, the integrandof(80), i.e. the function (78), is finite almost everywhere. Thus (80) means that ϕ is an L-integrable function, which means, again by (80), that $\mu_1 * \mu_2$ is absolutely continuous. Finally, if the Fundamental Theorem of Calculus is applied to (80), it follows that the derivative of $\mu_1 * \mu_2$, that is, its density, is represented by (78) for almost all x.

If both μ_1 and μ_2 are absolutely continuous, and if $\phi_1, \phi_2, \phi_{12}$ denote the densities of μ_1, μ_2, μ_{12} respectively, then (80) can be written in the symmetric form

$$(81) \qquad \phi_{12}(x) = \int \phi_1(x - u)\, \phi_2(u)\, du.$$

Convergent Convolutions

26. If $\mu_1, \dots, \mu_n$ are n distributions, (52) shows that $\mu_1 * \dots * \mu_n$ is a unique distribution; it has the transform

$$(82) \qquad (\mu_1 * \dots * \mu_n)^{\circ}(y) = \mu_1^{\circ}(y) \dots \mu_n^{\circ}(y)$$

and the squared standard deviation

$$(83) \qquad [\mu_1 * \dots * \mu_n]^2 = [\mu_1]^2 + \dots + [\mu_n]^2$$

($\leqq \infty$), as seen from (53) and from (67) by complete induction. Similarly, from (70) and (72),

$$(84) \qquad P(\mu_1 * \dots * \mu_n) = P(\mu_1) (+) \dots (+) P(\mu_n),$$

while, from (71),

$$(85) \qquad S(\mu_1 * \dots * \mu_n) = \text{closure of } S(\mu_1)\ (+) \dots (+)\ S(\mu_n).$$

Correspondingly, if $c_1, \dots, c_n$ are real numbers, then, from (51), (52) and (55),

$$(86) \qquad {}^{c_1}\mu_1 * \dots * {}^{c_n}\mu_n = {}^{c_1 + \dots + c_n}\varepsilon * \mu_1 * \dots * \mu_n,$$

since (15), (49) and the definition ${}^{c}\mu(x) = \mu(x - c)$ imply that

$$(87) \qquad {}^{a}\varepsilon * {}^{b}\varepsilon = {}^{a+b}\varepsilon.$$

In particular, if $(\mu_1)_1, \dots, (\mu_n)_1$ all exist, then, according to (63) and (37),

$$(88) \qquad ({}^{c_1}\mu_1 * \dots * {}^{c_n}\mu_n)_1 = \sum_{k=1}^{n} c_k + \sum_{k=1}^{n} (\mu_k)_1,$$

and so, by (38) or (64),

$$(89) \qquad (\mu_1 * \dots * \mu_n)^{\bullet} = \mu_1^{\bullet} * \dots * \mu_n^{\bullet},$$

where the existence of the moment on the left of (88), and therefore the existence of the distribution on the left of (89), is part of the statement.

It will often be necessary to consider the total variation of the distributions $\mu = \mu_1, \mu_2, \dots, \mu_n$ on the exterior of a fixed interval $-r \leqq x \leqq r$, that is, the integral on the left of (22). As $n \rightarrow \infty$, the numerical value of r turns out to be immaterial for the purposes at hand, and it will therefore be chosen to be 1. As an abbreviation, put

$$(90) \qquad \mu = (\int_{-\infty}^{\infty} - \int_{-1}^{1} d\mu; \text{ so that } \mu = 1 - \int_{-1}^{1} d\mu(x) \geqq 0,$$

by (1) and (3). It is easy to see that the symbol (90) cannot have any explicit property of the type (83).

27. For sake of shortness, let always

$$(91) \qquad \sum = \sum_{1}^{\infty} , \prod = \prod_{1}^{\infty}, \qquad \text{while } \int = \int_{-\infty} .$$

If $g_1, g_2, \dots$ is any sequence, the symbol $\sum g_n$ has two different meanings. On the one hand, it denotes just the ordered collection g_1, $g_1 + g_2, \dots$, which may or may not tend to a limit, and, on the other hand, it denotes the value represented by the limit whenever the latter exists. Correspondingly, if $\mu_1, \mu_2, \dots$ is a sequence of distributions, let the infinite convolution $\mu_1 * \mu_2 * \dots$ mean, on the one hand, the ordered collection of all distributions $\mu_1 * \dots * \mu_n$, which may or may not tend to a limit distribution, and, on the other hand, let it denote the limit distribution whenever the latter exists, i.e., whenever there exists a distribution μ satisfying $\mu_1 * \dots * \mu_n \to \mu$ as $n \to \infty$; in which case the infinite convolution will be called convergent (and in the contrary case divergent). Corresponding to (91), let $*\mu_n$ denote the infinite convolution $\mu_1 * \mu_2 * \dots$ in either case. Needless to say, this notation involves for the symbol $*$ two logically different meanings, as exemplified by the formal identity $\mu_0 * (*\mu_n) = *\mu_{n-1}$.

By a convergent convolution will be meant a convergent infinite convolution.

27. If μ_n and λ_n in (56) are replaced by $\mu_1 * \dots * \mu_n$ and $\lambda_1 * \dots * \lambda_n$ respectively, it is seen that

$$(*\mu_n) * (*\lambda_n) = *(\mu_n * \lambda_n) \tag{92}$$

holds for any pair of convergent convolutions $*\mu_n$, $*\lambda_n$, where the convergence of the convolution on the right of (92) is part of the statement.

It is clear from (15) and from the definition ${}^{a}\varepsilon(x) = \varepsilon(x - a)$ that $a_n \to a$ holds if and only if the distribution ${}^{a_n}\varepsilon$ tends to a limit distribution; in which case the latter is ${}^{a}\varepsilon$. It follows therefore by a two-fold application of both (86) and of the criterion (92), that, if $*\mu_n$ is a convergent convolution, the infinite convolution $*\,{}^{c_n}\mu_n$ is convergent if and only if the series $\sum c_n$ is; in which case

$$*\,{}^{c_n}\mu_n = {}^{c}\varepsilon * (*\mu_n), \text{ where } c = \sum c_n. \tag{93}$$

Another consequence of the criterion (92) is that, if $*\mu_n = \mu_1 * \mu_2 * \dots$ is a convergent convolution, the infinite convolution $*\mu_{n+m} = \mu_{m+1} * \mu_{n+2} * \dots$ is convergent for every fixed m and satisfies the relation

$$(94) \qquad (\mu_1 * \dots * \mu_m) * (*\mu_{n+m}) = *\mu_n .$$

This is seen from (51) and (52), if it is observed that, according to the case c = 0 of (55), the unit distribution ε satisfies $\mu * \varepsilon = \mu$, $\mu * \varepsilon * \varepsilon = \mu, \dots$ for every μ, and that every finite convolution can accordingly be thought of as a convergent infinite convolution.

If μ_n is replaced by $\mu_1 * \dots * \mu_n$ in the continuity theorem at the end of § 10, it is seen from (82) that an infinite convolution $*\mu_n$ is convergent if and only if there exists a continuous function to which the product on the right of (82) tends, for $-\infty < y < \infty$, as $n \longrightarrow \infty$; in which case the transform of the distribution $*\mu_n$ is

$$(95) \qquad (*\mu_n)^{\circ}(y) = \prod \mu_n^{\circ}(y),$$

and (82) tends to (95) uniformly on every fixed bounded y-interval.

In this criterion, the product on the right of (95) cannot beforehand be interpreted as uniformly convergent, or for that matter convergent, in the sense of the theory of infinite products $\prod g_n$. In fact, the latter theory declares $\prod g_n$ to be convergent only if there exists an N such that the number g_n does not vanish for any $n > N$ and the partial products of $\prod g_{n+N}$ tend to a non-vanishing limit as $n \longrightarrow \infty$; and both of these conditions imposed on N are required to be satisfied by a constant N, if $\prod g_n$ is called a uniformly convergent product of functions g_n. On the other hand, the above criterion refers merely to the uniform convergence of the partial products of the infinite product on the right of (95). However, these two interpretations of the uniform convergence of an infinite product, interpretations which clearly are inequivalent in general, prove to be equivalent in the present case.

In fact, suppose that (82) tends to (95) uniformly on every fixed bounded y-interval; so that the convergence of the infinite product (95) is not required in the sense of the theory of infinite products. Since (95) represents a transform, it is, by § 10, a continuous function which, by (12), cannot vanish for $-q < y < q$, if q is sufficiently small. Consequently, none of the factors on the right of (95) vanishes for $-q < y < q$, and so the infinite product (95) is uniformly convergent for $-q < y < q$, if uniform convergence is meant in the sense of the theory of infinite products. It follows that $\mu_n^{o}(y) \longrightarrow 1$ as $n \longrightarrow \infty$ holds uniformly for $-q < y < q$, if q is sufficiently small. But this implies, by the end of § 20, that $\mu_n^{o}(y) \longrightarrow 1$ as $y \longrightarrow \infty$ holds uniformly on fixed bounded y-interval. Hence, there exists for every bounded interval an N such that $\mu_n^{o}(y)$ does not vanish on the interval whenever $n > N$. This proves that the two interpretations of the uniform convergence of an infinite product are equivalent in the present case.

As an application, it is easy to show that, corresponding to Cauchy's criterion for the convergence of an infinite series $\sum g_n$, an infinite convolution $*\mu_n$ is convergent if and only if

(96) $$\mu_{n+1} * \dots * \mu_{n+k(n)} \longrightarrow \varepsilon, \text{as } n \to \infty, \text{ whenever } k = k(n) > 0$$

(that is, whenever k is a non-negative function of n).

In fact, (52) and (53) imply that the transform of the finite convolution on the left of (96) is the product of those factors of the infinite product (95) which have a subscript greater than n and not greater than n + k(n). Hence, if the distribution μ_n occurring in the continuity theorem at the end of §10 is identified with the finite convolution on the right of (96), it is seen from (15) that (96) holds if and only if the infinite product (95) is convergent on every fixed bounded y-interval. Since this was seen to be equivalent to the convergence of $*\mu_n$, it follows that (96) is necessary and sufficient for the convergence of $*\mu_n$.

It is understood that the finite convolution(96) is meant to denote the distribution μ_{n+1}, if k = k(n) is 0. Hence, a necessary condition for the convergence of $*\mu_n$ is represented by

(97) $$\mu_n \longrightarrow \varepsilon.$$

28. It will now be shown that, if $*\mu_n$ is any convergent convolution, a point x = s is in the spectrum of the distribution $*\mu_n$ if and only if there exists to x and to every n at least one point $s_n = s_n(x)$ in the spectrum of μ_n such that x = s appears in the form of the convergent series

$$s = \sum s_n. \tag{98}$$

(This does not claim, and it is not in general true, that $\sum s_n$ is a convergent series whenever s_n is in the spectrum of μ_n). It is clear from the definitions of A (+) B and lim A_n in § 24, that the assertion of the rule (98) can be formulated as follows: If $*\mu_n$ is a convergent convolution, then

$$S(*\mu_n) = \lim A_n, \text{ where } A_n = S(\mu_1)(+) \dots (+) S(\mu_n). \tag{99}$$

According to § 4, every spectrum is a closed set. On the other hand, the definition of lim A_n, formulated after (73), obviously implies that lim A_n is a closed set. But, if μ_n and μ in (73) are replaced by $\mu_1 * \dots * \mu_n$ and $*\mu_n$ respectively, it is seen that the limit of the set (85) contains the set $S(*\mu_n)$. Consequently, (99) will be proved if it is shown that the limit of the set (85) is contained in the set $S(*\mu_n)$.

The proof will be based on the fact that, if μ and λ is any pair of distributions, then the inequality

$$\int_{|x-c|<r} d\mu * \lambda(x) \geqq \int_{|v|<\frac{1}{2}r} d\mu(v) . \int_{|u-c|<\frac{1}{2}r} d\lambda(u) \tag{100}$$

holds for every point c and for every $r > 0$. For a moment, grant the truth of (100).

Suppose that $*\mu_n$ is convergent. Then (94) is valid (for every m). It implies, in view of (57), that $*\mu_{n+m} \rightarrow \varepsilon$ as $m \rightarrow \infty$. According to (15), this means that the total variation of $*\mu_{n+m}(x)$ on any fixed interval $-\frac{1}{2}r < x < \frac{1}{2}r$ tends to 1, as $m \longrightarrow \infty$, and exceeds, therefore, $\frac{1}{2}$ for every $m > N$, if $N = N_r$ is suitably chosen. Consequently, if (100) is applied to $\mu = \mu_1 * \dots * \mu_m$ and $\lambda = *\mu_{n+m}$, it follows from (94) that

$$\int_{|x-c|<r} d{*}\mu_n(x) \geqq \tfrac{1}{2} \int_{|x-c|<\frac{1}{2}r} d\,\mu_1*\dots*\mu_m(x) \text{ when } m > N_r,$$

where c and r are arbitrary. Hence it is clear from the definition of lim A_n after (73) and from the definition of $S(\mu)$ at the beginning of §4, that, if c is so chosen as to be contained in lim A_m, where $A_m = S(\mu_1*\dots*\mu_m)$, then c must be contained in $S({*}\mu_n)$ also. But this is precisely that assertion of (99) which remained to be proved.

In order to verify the general inequality (100), apply (49) to $\mu_1 = \mu$, $\mu_2 = \lambda$. It is then seen, from Fubini's theorem, that the integral on the left of (100) is identical with

$$\int\Big\{\int_{|x-c|<r} d_x\mu(x-u)\Big\}d\lambda(u), \text{ i.e., } \int\Big\{\int_{|v|<r} d_v\mu(v+c-u)\Big\}d\lambda(u).$$

Since $\int = \int_{-\infty}^{\infty}$, this repeated integral is not less than

$$\int_{|u-c|<\frac{1}{2}r}\Big\{\int_{|v|<r} d_v\mu(v+c-u)\Big\}d\,\lambda(u),$$

(the contribution of the omitted range $|u-c| \geqq \frac{1}{2}r$ being non-negative). Hence, in order to prove (100), it is sufficient to ascertain that, if u has any fixed value satisfying $|u-c| < \frac{1}{2}r$, then the interior integral in the last formula line is not less than the first integral on the right of (100). But this is clear from (1) if it is observed that, since u is restricted to the open interval having the length r and the midpoint c, the open interval having the length 2r and the midpoint c − u contains the open interval having the length r and the midpoint 0.

29. The necessary and sufficient conditions of § 27 for the convergence of $*\mu_n$ involve convergence properties of the functions occurring in (95) or in (96). It is therefore of interest that there are cases in which the convergence of $*\mu_n$ is ensured by the convergence of certain series involving moments of the first and second order only; so that that convergence properties of sequences of functions become replaced by corresponding properties of readily available sequences of numbers.

In this direction, it will be convenient to consider first the convolutions $*\mu_n$ in which every μ_n is centered in the sense of (38). In this case (to which the general case will subsequently be reduced), a sufficient condition for the convergence of $*\mu_n$ can be formulated as follows:

(101) If $\sum [\mu_n]^2 < \infty$, and if every μ_n is centered, then $*\mu_n$ is convergent; moreover, $*\mu_n$ is a centered distribution, and $[*\mu_n]^2$ is exactly $\sum [\mu_n]^2$.

Since $\sum [\mu_n]^2 < \infty$ implies that $[\mu_n] < \infty$ for every n, it is clear that (101) is not a necessary condition for the convergence of a convolution of centered distributions. However, the condition $\sum [\mu_n]^2 < \infty$ turns out to be necessary for convergence in a particular case of considerable interest. In fact, this will be the case if the spectra of $\mu_1, \mu_2, \ldots$ are uniformly bounded, i.e., if $S(\mu_n)$ is contained in a fixed bounded interval for every n. Thus the assertions of (101) can, in this case, be completed as follows:

(102) If $\mu_1, \mu_2, \ldots$ are centered distributions with uniformly bounded spectra, then the convolution $*\mu_n$ is convergent if and only if the numerical series $\sum [\mu_n]^2$ is.

In order to prove (101), notice first that, since $\sum [\mu_n]^2 < \infty$ implies that $[\mu_n] < \infty$, it is the second line of (41) that applies to every n. Furthermore, (101) assumes (38) for every n. Accordingly, $[\mu_n]^2 = (\mu_n)_2$, since $(\mu_n)_1 = 0$. Thus (32) is applicable to $\mu = \mu_n$, if $k = 2$ and, therefore, if $k = 1$; so that $D_y^2 \mu_n^o(0) = -[\mu_n]^2$ and $D_y^1 \mu_n^o(0) = 0$. It follows therefore from (12) and from the case $k = 2$ of (31) that, in view of Taylor's formula,

$$\mu_n^o(y) = 1 - \tfrac{1}{2}[\mu_n]^2 y^2 - \tfrac{1}{2}y^2 \int x^2 (\cos(\theta'_{ny}\, xy) + i \sin(\theta''_{ny}\, xy))\, d\mu_n(x),$$

where the θ are independent of x and satisfy the inequalities $0 < \theta < 1$. In particular, $|\mu_n^o(y) - 1| < 2[\mu_n]^2 y^2$. Since $\sum [\mu_n]^2$ is a convergent series of non-negative constants, this inequality implies that the infinite product on the right of (95) is uniformly convergent on every fixed bounded y-interval. In view of §27, this proves the first assertion of (101), namely the convergence of $*\mu_n$.

The convergence of $\sum [\mu_n]^2$ and the preceding inequality for the deviation of $\mu_n{}^o(y)$ from 1 also imply that there exists an interval $-a < y < a$ such that the finite product on the right of (82) deviates from 1 by not more than Cy^2 if $-a < y < a$, where both constants a, C are independent of the index n occurring in (82). It follows therefore from (95) that $|(*\mu_n)^o(y) - 1| < Cy^2$ for $-a < y < a$. In other words, the case $n = 1$ of (34) is satisfied by $c_0 = 1$ and $c_1 = 0$, if $\mu = *\mu_n$. It follows therefore from §12 that there exists a constant c_2 satisfying (33') for $k = 2$ and that $(\mu)_2 < \infty$, finally that (33) is applicable to $k = 2$. But (33') means, in the present case, that $\mu^o(y) = 1 + c_2y^2 + o(y^2)$ as $y \rightarrow 0$. If this is compared with the case $k = 2$ of (33), it follows that $c_2 = -\frac{1}{2}(\mu)_2$. Accordingly, the distribution $\mu = *\mu_n$ has a finite second moment satisfying $(*\mu_n)^o = 1 - \frac{1}{2}(*\mu_n)_2 y^2 + o(y^2)$. This, when compared again with (33), implies that $(*\mu_n)_1 = 0$. In view of the definition (38), this means that $(*\mu_n)$ is centered. This is the second assertion of (101).

Since $(*\mu_n)_2 < \infty$ and $(*\mu_n) = 0$, the second line of (41) gives $[*\mu_n]^2 = (*\mu_n)_2$. If this is substituted into the o-relation just found, it is seen that the last assertion of (101) is equivalent to the statement that the deviation of the product (95) from 1 is $-\frac{1}{2}y^2 \sum [\mu_n]^2 + o(y^2)$ as $y \rightarrow 0$. But the truth of this statement is readily verified from the last formula line, since $\int x^2 d\mu_n(x) = [\mu_n]^2$ and $\sum [\mu_n]^2 < \infty$.

In order to prove (102), let $*\mu_n$ be a convergent convolution of centered distributions (for the present, the spectra $S(\mu_n)$ need not be uniformly bounded). Put $\lambda_n = {}_{-1}\mu_n * \mu_n$, where ${}_{-1}\mu_n(x)$ denotes, as before (17), the distribution represented by $1 - \mu_n(-x)$ at its continuity points. Thus it is clear that ${}_{-1}\mu_n$ is centered and that $\left[{}_{-1}\mu_n\right] = \left[\mu_n\right]$. It follows therefore from (64) and (67) that $\lambda_n = {}_{-1}\mu_n * \mu_n$ is centered and that $\left[\lambda_n\right]^2 = 2\left[\mu_n\right]^2$.

Furthermore, $\lambda_n^{\circ}(y) = |\mu_n^{\circ}(y)|^2$, by (54), and so a repeated application of the product criterion of § 27 shows that the convergence of $*\mu_n$ implies, on the one hand, the convergence of $*\lambda_n$ and, on the other hand, the uniform convergence of $\prod \lambda_n^{\circ}(y)$ on any fixed bounded y-interval. But $\lambda_n^{\circ}(y)$ is non-negative, hence real, and so it is clear from the case $\mu = \lambda_n$ of (11) that $\prod \lambda_n^{\circ}(y)$ must be absolutely convergent because of its convergence. Consequently, $\sum |1 - \lambda_n^{\circ}(y)| < \infty$ for every y, which implies, by the case $\mu = \lambda_n$ of (21), that

$$\int_{-r}^{r} x^2 d\lambda_n(x) < \infty \text{ for every } r.$$

Suppose now that the spectra $S(\mu_n)$ are uniformly bounded. Then it is clear that the same is true of the "reflected" spectra $S({}_{-1}\mu_n)$, and so, by (71), of the spectra $S({}_{-1}\mu_n * \mu_n)$. It follows therefore from $\lambda_n = {}_{-1}\mu_n * \mu_n$ that there exists an r such that each of the distributions $\lambda_n(x)$ is constant for $-\infty < x < -r$ and for $r < x < \infty$. Hence, the last formula line means that $\sum (\lambda_n)_2 < \infty$. But λ_n is centered, i.e., $(\lambda_n)_2 = \left[\lambda_n\right]^2$; so that $\sum \left[\lambda_n\right]^2 < \infty$.

This completes the proof of (102), since, as observed before, $\left[\lambda_n\right]^2 = 2\left[\mu_n\right]^2$.

30. It is easy to transcribe the results of §29 to the case of distributions μ_n which are not centered. In fact, (101) can be generalized as follows:

(103) If $\sum [\mu_n]^2 < \infty$, then the convolution $*\mu_n$ is convergent if and only if the series $\sum (\mu_n)_1$ is; in which case $[*\mu_n]^2 = \sum [\mu_n]^2$ and $(*\mu_n)_1 = (\mu_n)_1$.

The corresponding transcription of (102) runs as follows:

(104) If $\mu_1, \mu_2, \ldots$ are distributions having uniformly bounded spectra, then the convolution $*\mu_n$ is convergent if and only if both series $\sum [\mu_n]^2$, $\sum (\mu_n)_1$ are.

In order to prove (103), suppose that $\sum [\mu_n]^2 < \infty$. Then $[\mu_n] < \infty$; hence $(\mu_n)_2 < \infty$, and so the first moment $(\mu_n)_1$ and the centered distribution $\mu_n^{\bullet}$ exist for every n. But the second line of (41) implies that $[\mu_n] = [\mu_n^{\bullet}]$, since $(\mu_n^{\bullet})_1 = 0$. Furthermore, if $c_n = (\mu_n)_1$, then (38) shows that ${}^{c_n}\mu_n = \mu_n^{\bullet}$ and $({}^{c_n}\mu_n)^{\bullet} = \mu_n^{\bullet}$. Hence, if (101) and the criterion of (93) are applied to the case in which μ_n is replaced by $\mu_n^{\bullet}$, and if (37) is then applied to $\mu = *\mu_n^{\bullet}$ and to $c = \sum c_n$, where $c_n = (\mu_n)_1$, it is seen that (103) is equivalent to its particular case (101).

Correspondingly, (104) follows from (102), since the spectra $S(\mu_n)$ are uniformly bounded if and only if the spectra $S(\mu_n^{\cdot})$ are. In fact, it is clear from (30), where $(\mu)_0 = 1$, that, if the spectrum of a distribution μ is contained in an interval, then the latter contains the point $x = (\mu)_1$; so that the assertion is clear from (38).

31. For every distribution μ, define another distribution, $\mu^{:}$, by assigning the transform $\mu^{:\circ} = \mu^{:\circ}(y)$ of $\mu^{:} = \mu^{:}(x)$ as the function

$$(105) \qquad \mu^{:\circ}(y) = \{\mu\} + \int_{-1}^{1} e^{ixy}\, d\mu(x), \quad \text{where } \{\mu\} = \text{const.}$$

If (12) is applied to both μ and $\mu^{:}$, it is seen from (10) that the constant occurring in (105) must be the non-negative constant (90). Correspondingly, comparison of (105) with (10) shows that the distribution $\mu^{:}$ is defined as follows:

$$(106) \qquad d\mu^{:}(x) = \begin{cases} 0 \text{ if } |x| > 1, \\ d\mu(x) \text{ if } 0 < |x| < 1, \\ \{\mu\} + d\mu(x) \text{ if } x = 0. \end{cases}$$

It is also seen from (105) and (10) that

$$(107) \qquad \mu^{:\circ}(y) - \mu^{\circ}(y) = \left(\int_{-\infty}^{\infty} - \int_{-1}^{1}\right)(1 - e^{ixy})\, d\mu(x),$$

which implies, by (90), that

$$(108) \qquad |\mu^{:\circ}(y) - \mu^{\circ}(y)| \leqq 2\{\mu\}.$$

Since the spectrum of $\mu^{:}$ is contained in the interval $-1 < x < 1$, it follows from (104) that, for any sequence of distributions $\mu_1, \mu_2, \ldots$,

(109) $*\mu_n^{:}$ is convergent if and only if both $\sum [\mu_n^{:}]^2$ and $\sum (\mu_n^{:})_1$ converge.

Thus there arises the question as to the possibility of a connection between the convergence of $*\mu_n$ and the convergence of $*\mu_n^{:}$, where $\mu_1, \mu_2, \ldots$ is an arbitrary sequence of distributions. In this direction, it is easy to prove that

(110) $\sum \{\mu_n\} < \infty$ implies that either both $*\mu_n$ and $*\mu_n^{:}$ are convergent or both are divergent.

In fact, if $z_1, z_2, \ldots$ and $w_1, w_2, \ldots$ are two sequences of numbers satisfying $z_n \longrightarrow 1$ and $w_n \longrightarrow 0$ respectively, then $\log (z_n + w_n) - \log z_n$, being equal to $\log (1 + w_n/z_n)$, is $O(|w_n|)$ as $n \rightarrow \infty$. Hence, if $\prod z_n$ and $\sum |w_n|$ are convergent, then $\prod (z_n + w_n)$ is convergent. Obviously, a corresponding statement holds uniformly in y for the case of functions $z_n = z_n(y)$, $w_n = w_n(y)$. Thus, if these functions are given either by $z_n = \mu_n^{o}$, $z_n + w_n = \mu_n^{:o}$ or by $z_n = \mu_n^{:o}$, $z_n + w_n = \mu_n^{o}$, it is seen from the product criterion of (95), that (110) follows from (108).

32. Let an infinite convolution $*\mu_n$ be called absolutely convergent if $*\lambda_n$ is convergent whenever $\lambda_1, \lambda_2, \ldots$ is a rearrangement of $\mu_1, \mu_2, \ldots$.

According to the criterion of (95), this will be the case if and only if the infinite product $\prod \lambda_n^o(y)$ belonging to an arbitrary rearrangement of $\mu_1, \mu_2, \ldots$ is uniformly convergent on every fixed bounded y-interval. But this is equivalent to the absolute and uniform convergence of the product (95) on any fixed bounded y-interval. This, when compared with (19) and with the fact that the absolute and the uncondition convergence of a product $\prod g_n$ are equivalent notions, implies that a convergent convolution $*\mu_n$ is absolutely convergent if and only if the distribution $*\lambda_n$ is identical with the distribution $*\mu_n$ whenever the convolution $*\lambda_n$ is a convergent rearrangement of the convolution $*\mu_n$.

Since the absolute convergence of $\prod g_n$ is equivalent to $\sum |1 - g_n| < \infty$ it also follows that $*\mu_n$ is absolutely convergent if and only if $\sum |1 - \mu_n^o(y)|$ is uniformly convergent on every fixed bounded y-interval.

If this interval is chosen to be $-1 \leqq y \leqq 1$, and if (22) is applied to $r = 1$ and to $\mu = \mu_n$, it follows, in particular, that the convergence of the non-negative numerical series $\sum \{\mu_n\}$, where $\{\mu\}$ is defined by (90), is a necessary condition for the absolute convergence of $*\mu_n$. Hence, it is clear from (110) that the absolute convergence of $*\mu_n$ implies the absolute convergence of $*\mu_n^{:}$, and that the absolute convergence of $*\mu_n^{:}$ implies the absolute convergence of $*\mu_n$ whenever $\sum \{\mu_n\} < \infty$. It follows therefore from (109) that the simultaneous satisfaction of the three conditions

$$(111) \quad \sum \{\mu_n\} < \infty, \quad \sum [\mu_n{}^{:}]^2 < \infty, \quad \sum |(\mu_n{}^{:})_1| < \infty$$

is necessary and sufficient for the absolute convergence of an infinite convolution $*\mu_n$.

If a convergent product $\prod g_n$ is written in the form $\prod (1 - h_n)$, and if $g_n \leqq 1$ for every n, then, since $h_n > 0$, it follows that $\prod g_n$ is absolutely convergent. If this remark is applied to $g_n = \mu_n{}^{\circ}(y)$, it is seen from the case $\mu = \mu_n$ of (11) that, if transforms $\mu_n{}^{\circ}(y)$ are real-valued for every y (which means that the distributions μ_n are symmetric; cf. § 67 below), then the convolution $*\mu_n$ cannot converge unless it converges absolutely.

33. It turns out that $\sum \{\mu_n\} < \infty$ is a necessary condition for the mere convergence of $*\mu_n$. However, this fact is substantially more recondite, that is, its particular case considered in § 32, where $*\mu_n$ is assumed to converge absolutely. In fact, it will be convenient to defer the proof to § 39, where use will be made of the theory of product measures on infinite product spaces, rather than of the theory of the transform (10).

If the necessity of $\sum\{\mu_n\} < \infty$ for the convergence of $*\mu_n$ has been proved, then (109) and (11) clearly imply the following "three-series theorem": An infinite convolution $*\mu_n$ is convergent if and only if all three numerical series

$$(112) \qquad \sum\{\mu_n\}, \quad \sum\left[\mu_n^{:}\right]^2, \quad \sum(\mu_n^{:})_1$$

are convergent. The easier criterion of § 32 for the absolute convergence of $*\mu_n$, namely the criterion (111), is of course a particular case. Needless to say, the terms of the first two of the three series (112) are non-negative for every $*\mu_n$.

The definitions (90) and (106) introduce into the three-series theorem the absolute constant $r = 1$. Actually, the theorem remains valid if $r = 1$ is replaced by any fixed $r > 0$ in (90), (106) and (112). In fact, (90) and (106) show that this replacement is equivalent to a change of the unit of length on the x-axis and therefore, in view of (17), to the reciprocal change of the unit of length on the y-axis. But the product (95) is uniformly convergent on every fixed bounded y-interval if and only if the same is true of the product $\prod \mu_n^{\circ}(ry)$ for every fixed $r > 0$. Since both this product criterion and the convergence of all three series (112) represent necessary and sufficient conditions for the convergence of $*\mu_n$, the assertion follows.

A curious corollary is that there exists for every convergent convolution $*\mu_n(x)$ a convergent series $\sum c_n$ such that the convolution $*\mu_n(x - c_n)$ is absolutely convergent. In order to see this, it is sufficient to place $c_n = -(\mu_n^{\cdot})_1$ and to compare the three-series theorem with the criteria of (111) and (93); cf. (38) and (40).

A consequence of this corollary is that, if $\mu = {*}\mu_n$ is a convergent convolution and if $\lambda = {*}\lambda_n$ in any of its convergent rearrangements (cf. § 32), then there exists to this rearrangement a constant c such that $\lambda = {}^{c}\mu$, i.e., $\lambda(x) = \mu(x - c)$. In order to see this, it is sufficient to choose convergent series $\sum a_n$, $\sum b_n$ which make the infinite convolutions ${*}\mu_n(x - a_n)$, ${*}\lambda_n(x - b_n)$ absolutely convergent, and then to apply the remarks leading to (92) and (93); cf. (86) and (87).

Convergence in Measure

34. Let T be a "space" of "points" t. Let there be given on T a field of sets F each of which consists of points t, and suppose that every point t and the space T are sets F (i.e., sets contained in the field). Suppose finally that there is given on the field a measure $\phi = \phi(F)$ for which the measure of T is 1. In other words, let there be given on the field a function $\phi(F)$ satisfying, corresponding to (1), (2) and (3), the conditions that $\phi(F) \geqq 0$ for every F, that $\phi(F_1) + \phi(F_2) + \dots = \phi(F_1 + F_2 + \dots)$ for every sequence of mutually disjoint sets $F_1, F_2, \dots$, finally that

$$\phi(T) = 1. \tag{113}$$

Let there be given on T a function $x(t)$ which has values represented by points of the line X: $-\infty < x < \infty$ and which, with reference to the given field on T, is a measurable function of the position t on T. It is understood that $x(t)$ need not be defined, or can be $\pm\infty$, on a t-set of ϕ-measure 0.

It is easy to see that the assignment

$$\mu_E = \phi(F_{x(t) \text{ in } E}) \tag{114}$$

defines on the field of the Borel sets E on the line X a set-function μ_E which is a distribution in the sense of § 1. It is understood that the symbol within the parenthesis on the right of (114) denotes the set of those points t of T for which $x(t)$ is a point x of E. Thus it is sufficient to show that, if E^x denotes, as at the beginning of § 5, the half-line consisting of those points of the line X which are to the left of the point x, then (114) defines for every $E = E^x$ a value which, when denoted by $\mu(x)$, satisfies (8). But the expression on the right of (114) is defined for every $E = E^x$, the function $x(t)$ having been supposed to be ϕ-measurable on T. And the three conditions (8) for μ follow from the three conditions assumed, before and by (113), for the measure ϕ on T, the function $x(t)$ having been assumed to represent a real number x ($\neq \pm \infty$) for almost all points t of T.

The distribution $\mu = \mu_E$ or $\mu = \mu(x)$ which is defined by (114) on X will be called the ϕ-distribution of $x(t)$. It is characterized by the property that, if $f(x)$ is any Baire function on X satisfying either $|f(x)| <$ const. or $f(x) \geqq 0$, then

$$(115) \qquad \int_X f(x)\, d\mu = \int_T f(x(t))\, d\phi.$$

In fact, (115) follows from (114) in either case, while (114) follows from (115) by placing $f(x) = 1$ or $f(x) = 0$ according as x is or is not in the Borel set E.

35. Let $x_1(t), x_2(t), \ldots$ be a sequence of functions on T such that every $x_n(t)$ satisfies, with reference to a fixed $\phi = \phi(F)$ on T, the same conditions as the function $x(t)$ in § 34.

If there exists for every constant $q > 0$ and for every constant $p > 0$ an $N = N_q^p$ such that the set of those points t of T at which $|x(t) - x_n(t)|$ exceeds p has a ϕ-measure less than q whenever $n > N$, then $x_n(t)$ is said to tend (on T) to $x(t)$ in ϕ-measure, as $n \rightarrow \infty$.

It is known that, if $x_n(t)$ tends to $x(t)$ in ϕ-measure, then the sequence $x_1(t)$, $x_2(t),\ldots$ contains a subsequence, say $x_{(1)}(t)$, $x_{(2)}(t),\ldots$, such that $x_{(n)}(t) \longrightarrow x(t)$ as $n \rightarrow \infty$ holds for every fixed t not contained in a certain t-set of ϕ-measure 0. This implies that, if $x_n(t)$ tends to $x'(t)$ and also to $x''(t)$ in ϕ-measure, then $x'(t) = x''(t)$ holds for almost all t (that is, for every t not contained in a certain t-set of ϕ-measure 0).

If the functions $x_n(t)$, instead of tending to a function $x(t)$ in ϕ-measure, have the property that there exists for every constant $q > 0$ and for every constant $p > 0$ an $N = N_q^p$ such that the set of those points t of T at which $|x_m(t) - x_n(t)|$ exceeds p has a ϕ-measure less than q whenever $m > N$ and $n > N$, then the sequence $x_1(t)$, $x_2(t),\ldots$ is said to be convergent (on T) in ϕ-measure.

It is obvious that every sequence $x_1(t)$, $x_2(t),\ldots$ which tends to a function $x(t)$ in ϕ-measure is convergent in ϕ-measure. The fundamental fact (compactness theorem) of this theory is known to be the converse of this obvious remark, namely the fact that every sequence $x_1(t)$, $x_2(t),\ldots$ which is convergent in ϕ-measure tends to a function $x(t)$ in ϕ-measure.

In order to establish a connection between these notions and the definition (114), suppose that $x_n(t)$ tends to $x(t)$ in ϕ-measure. This means that the set of those points t of T at which $|x_n(t) - x(t)|$ exceeds a fixed $\varepsilon > 0$ tends to 0 as $n \longrightarrow \infty$. Hence, if μ_n and μ denote the ϕ-distributions of $x_n(t)$ and of $x(t)$ respectively, then, since ϕ is fixed, it is readily seen from the definition of a ϕ-distribution (§ 34), that $\mu_n \longrightarrow \mu$ holds in the sense defined at the end of § 2. This, when combined with the compactness theorem quoted before, proves the following fact:

(116) If a sequence $x_1(t)$, $x_2(t)$, ... is convergent (on T) in ϕ-measure, and if μ_n denotes the distribution on X which represents the ϕ-distribution of $x_n(t)$, then there exists on X a distribution μ satisfying $\mu_n \longrightarrow \mu$; in fact, this μ is precisely the ϕ-distribution of the function $x(t)$ to which $x_n(t)$ tends (on T) in ϕ-measure.

However, (116) supplies only a necessary, and not a sufficient, condition for convergence in ϕ-measure. In order to see this, let T be the interval $-\frac{1}{2} \leqq t \leqq \frac{1}{2}$ and ϕ the Euclidean measure on it, finally $x_n(t) = (-1)^n \operatorname{sgn} t$. Then $|x_n(t) - x_{n+1}(t)| = 2$ holds for every n and for every $t \neq 0$, and so the sequence $x_1(t)$, $x_2(t)$,... cannot converge on T in ϕ-measure. Nevertheless, there exists a distribution μ satisfying $\mu_n \longrightarrow \mu$, since μ_n is independent of n (in fact, $\mu_n(x)$ is 0, $\frac{1}{2}$ or 1 according as $x < -1$, $|x| < 1$ or $x > 1$).

In view of this failure of the converse of (116), it is a fundamental property of the particular case of infinite produce measures, to be defined in § 37, that the converse of (116) will, in § 38, turn out to be correct for this particular case.

36. The meaning of the convolution of distributions μ on X can constructively be described in terms of the notions of § 34, if use is made of the idea of product measures.

In order to see this, let $T_1, T_2, \ldots$ be spaces of respective points

$t_1, t_2, \ldots$ and let every T_n carry a measure $\phi_n = \phi_n(F_n)$ on a field of subsets F_n of T_n in such a way that all the conditions required before and by (113) in §34 are satisfied for every n; in particular,

(117) $$\phi_n(T_n) = 1.$$

For a fixed n, let $F^n = F_1 . F_2 . \ldots . F_n$ denote the "Cartesian product" of sets $F_1, \ldots, F_n$ which are chosen in the respective fields on $T_1, T_2, \ldots, T_n$. Thus,

(118) $$T^n = T_1 . \ldots . T_n$$

denotes the space in which an arbitrary point

(119) $$t^n = t_1 . \ldots . t_n$$

has the coordinates $t_1, \ldots, t_n$; coordinates which are arbitrary points of $T_1, \ldots, T_n$ respectively.

Consider on the product space T^n the "product measure" $\phi^n = \phi^n(F^n)$, defined by placing

(120) $$\phi^n(F^n) = \phi_1(F_1)\ \phi(F_2)\ \ldots\ \phi_n(F_n) \text{ if } F^n = F_1 . F_2 . \ldots . F_n .$$

The collection of those sets in T^n which are representable as product sets $F_1 . \ldots . F_n$ does not in general form a field. However, this collection can always be extended so as to become a field and to possess the extremal property that, if a set of points (119) is in the field and has the form $G . F_2 . F_3 . \ldots . F_n$, where G is any set of points t_1 on T_1, then G is a set F_1 contained in the given field on T_1; and that the same holds if T_1 is replaced by $T_2, \ldots, T_n$. In the sequel, F^n will denote any set contained in the resulting field of sets of points (119). It can be shown that the assignment (120) defines on the field of all these sets F^n a unique measure

ϕ^n. Finally, since (119) holds for n = 1,2,..., it is clear from (117), (118) and (120) that, corresponding to (113),

$$(121) \qquad \phi^n(T^n) = 1.$$

For every n, let there be given on T_n a ϕ_n-measurable function $x_n(t_n)$ of the position t_n, where it is understood that $x_n(t_n)$ need not be defined, or can be $\pm\infty$, on a t_n-set of ϕ_n-measure 0. Then, in virtue of (119), the function represented by the sum

$$(122) \qquad x^n(t^n) = x_1(t_1) + \dots + x_n(t_n)$$

is a function of the position t^n on the product space T^n. Clearly, the function $x^n(t^n)$ has, with reference to the product measure ϕ^n on T^n, all the properties required of $x_n(t_n)$ on T_n with reference to ϕ_n. Hence, there exists on X a distribution, say μ^n, representing the ϕ^n-distribution of $x^n(t^n)$ on T^n, as defined in §34. It will now be shown that, if μ_n denotes, for n = 1,2,..., that distribution on X which is the ϕ_n-distribution of $x_n(t_n)$ on T_n, then

$$((123) \qquad \mu^n = \mu_1 * \dots * \mu_n.$$

In fact, let T, t, ϕ, x(t) in (115) be respectively identified with (118), (119), (120), (122), and therefore μ with the ϕ^n-distribution of $x^n(t^n)$ on T^n, i.e., with μ^n. Finally, let the function f(x) occurring in (115) be chosen to be 1 or 0 according as x is or is not in E, where E is a fixed Borel set on X. Then it is clear that (115) becomes

$$(124) \qquad \mu^n_E = \int \dots \int_{x_1(t_1)+\dots+x_n(t_n) \text{ in } E} d\phi_1 \dots d\phi_n,$$

where the n-fold integration is extended over the set of those points (119) of (118) which satisfy the restriction specified under the integral signs.

Now let (115) be applied to each of the functions $x(t) = x_1(t_1),\ldots,x_n(t_n)$ of the positions $t_1,\ldots,t_n$ on the respective factor spaces $T_1,\ldots,T_n$ and to the respective measures $\phi = \phi_1,\ldots,\phi_n$ on these spaces. Then it is seen from (12) and from Fubini's theorem that (124) can be written in the form

$$(125) \qquad \mu^n{}_E = \int\cdots\int_{x_1+\ldots+x_n \text{ in } E} d\mu_1(x_1)\ \ldots\ d\mu_n(x_n).$$

But (48) shows that (125) is identical with the definition of $\mu_1 * \ldots * \mu_n$ for $n = 2$ and therefore, as observed after (52), for every n. This proves (123).

It should be noted for later reference that, if r is any positive constant,

$$(126) \quad \int_{|x|>r} d\mu_1 * \ldots * \mu_n(x) = \int\cdots\int_{|x_1+\ldots+x_n|>r} d\mu_1(x_1)\ \ldots\ d\mu_n(x_n) =$$

$$\phi_1\phi_2\ldots\phi_n\text{-measure of the } (t_1,\ldots,t_n) \text{set on which } |x_1(t_1)+\ldots+x_n(t_n)| > r$$

This follows by applying (125) and (124) to the Borel set E consisting of the two half-lines exterior to the interval $-r \leqq x \leqq r$.

37. Let

$$(127) \qquad T = T_1.T_2.\ \ldots$$

and

(128) $$t = t_1.t_2. \dots$$

denote what result by letting $n \longrightarrow \infty$ in (117) and (118); so that t is a point of the infinite product space. Correspondingly, let F denote any set in a certain field of sets in T; a field which contains every infinite product set $F_1.F_2. \dots$, where F_n is a set in the field on T_n. It is known that, due to (121), it is possible to define on T a measure $\phi = \phi(F)$ by placing, corresponding to (120),

(129) $$\phi(F) = \phi(F_1)\ \phi(F_2) \dots \text{ if } F = F_1.F_2. \dots ;$$

so that, in particular,

(130) $$\phi(T) = 1,$$

by (127) and (117).

The function (122) of the position (119) on (118) can be thought of as a function $x^n(t)$ of the position (128) on (127), since the assignment

(131) $$x^n(t) = x_1(t_1)+\dots+x_n(t_n), \text{ where } t = t_1.t_2. \dots,$$

defines on the infinite product space a single-valued function; a function satisfying

(132) $$x^n(t') = x^n(t'') \text{ whenever } t'_m = t''_m \text{ for every } m \leq n.$$

In particular, the n-th term of (131), a term which in § 36 was given as a function of the position t_n on the n-th factor space T_n, becomes a function $x_n(t)$ of the position on the infinite product space T:

(133) $$x_n(t) = x_n(t_n), \text{ where } t = t_1.t_2. \dots,$$

it being understood that, corresponding to (132),

(134) $$x_n(t') = x_n(t'') \text{ whenever } t'_n = t''_n.$$

In §36, the distributions μ_n and μ^n on X were defined to be the ϕ_n-distribution of $x_n(t_n)$ and the ϕ^n-distribution of $x^n(t^n)$ respectively. In view of (133), (134), (129), (130) and (117), the first of these definitions is seen to be equivalent to the assignment

(135) $\mu_n = \phi$-distribution of $x_n(t)$,

and the second, by (132) and (131), to

(136) $\mu^n = \phi$-distribution of $x^n(t)$,

where $\phi = \phi_1 \phi_2 . \ldots$ and $t = t_1 . t_2 . \ldots$ in both cases. In particular, the second line of (126) can be written in the following form:

(137) ϕ-measure of the t-set on which $x_1(t) + \ldots + x_n(t) > r$.

In view of the result (123) of §36, where $n < \infty$, it is worth emphasizing that all of the above remarks involving (127), (128) and (129), where $n = \infty$, hold for any sequence of measures $\phi_1, \phi_2, \ldots$ satisfying (117); so that the resulting distributions (135) on X need not be such as to make the infinite convolution $*\mu_n$ convergent, i.e., such as to lead to an extension of (123) to the present case, $n = \infty$. Correspondingly, the series $\sum x_n(t)$, in which the terms $x_n(t)$ are defined by (133) and (134) as functions of the position (128) on the infinite product space (127), has an n-th partial sum, $x^n(t)$, which in no sense need tend to a function , say $x(t)$, as $n \longrightarrow \infty$; in fact, the terms of this partial sum, which is the sum (131), are allowed to be arbitrary within the trivial limitations specified before (122).

38. It will now be shown that the question of whether or not the series $\sum x_n(t)$ defines on the infinite product space T a function, say x(t), depends on the convergence or divergence of the infinite convolution $*\mu_n$. An appropriate way of defining an x(t) turns out to be the convergence of $\sum x_n(t)$ in ϕ-measure, where ϕ is the infinite product measure (129) on the infinite product space T of the points $t = t_1.t_2. \dots$. It is understood that $\sum x_n(t)$ is said to tend to x(t) in ϕ-measure (on T) if the n-th partial sum of $\sum x_n(t)$, that is, the function $x^n(t)$ on T, tends to x(t) in ϕ-measure (on T); cf. § 35. Then the connection just mentioned can be formulated as follows:

(138) If the distribution $\mu_n = \mu_n(x)$, where $-\infty < x < \infty$ and $n = 1,2,\dots$, is the ϕ_n-distribution of the function $x_n(t_n)$ of the position t_n on T_n, and if ϕ denotes the infinite product measure $\phi_1 \phi_2 \dots$ on the space $T = T_1.T_2. \dots$ of the points $t = t_1.t_2. \dots$, then the series $\sum x_n(t)$, where $x_n(t)$ is defined by (133), is convergent (on T) in ϕ-measure if and only if the infinite convolution $*\mu_n$ is a convergent infinite convolution; in which case the distribution $*\mu_n$ on X is precisely the ϕ-distribution of the function x(t) to which the series $\sum x_n(t)$ converges (on T) in ϕ-measure.

In order to prove (138), suppose that $*\mu_n$ is convergent. The definition (15) and the necessary and sufficient criterion (96) show that this will be the case if and only if the total variation of the distribution $\mu_{n+1} * \dots * \mu_{n+k}$ on the two half-lines exterior to any fixed interval $-r \leqq x \leqq r$ tends to 0 as $n \to \infty$, where $k = k(n)$ is any non-negative function of n. It is clear from (48) and from the remarks following (52), that this total variation is represented by the integral of $d\mu_{n+1}(x_{n+1}) \dots d\mu_{n+k}(x_{n+k})$, extended over the manifold of those points of the k-dimensional Euclidean $(x_{n+1}, \dots, x_{n+k})$-space which satisfy the inequality $|x_{n+1} + \dots + x_{n+k}| > r$. Hence, if the n indices $1, \dots, n$ occurring in the identity (126) and in the identical transcription (137) of (126) are replaced by the k indices $n + 1, \dots, n + k$, it follows that the total variation of the distribution $\mu_{n+1} * \dots * \mu_{n+k}$ on the two half-lines exterior to any fixed interval $-r \leqq x \leqq r$ is identical with the ϕ-measure of the set of those points t of T which satisfy the inequality $|x_{n+1}(t) + \dots + x_{n+k}(t)| > r$. This inequality can be written in the form $|\overset{n+k}{x}(t) - \overset{n}{x}(t)| > r$, since $\overset{n}{x}(t)$ denotes the n-th partial sum of the series $\sum x_n(t)$.

Consequently, the infinite convolution $*\mu_n$ is convergent if and only if the ϕ-measure of those points t which satisfy the inequality $|\overset{n+k}{x}(t) - \overset{n}{x}(t)| > r$ tends to 0 as $n \longrightarrow \infty$, if $r > 0$ is arbitrarily fixed and $k = k(n)$ is any non-negative function of n. But this means, by § 35, the convergence (on T) in ϕ-measure of the sequence $\overset{1}{x}(t), \overset{2}{x}(t), \dots$ on T. Since this sequence is the sequence of the partial sums of the series $\sum x_n(t)$, the proof of the necessary and sufficient criterion, asserted by the first part of (138), is complete.

In order to prove the remaining assertion of (138), suppose that $*\mu_n$ is convergent and denote by $x(t)$ the function to which the partial sums, $\overset{n}{x}(t)$, of $\sum x_n(t)$ tend on T in ϕ-measure. Then, if μ and μ^n denote the ϕ-distributions of $x(t)$ and $\overset{n}{x}(t)$ respectively, it follows from the last of the assertions of (116), that $\mu^n \longrightarrow \mu$. This means, by (136) and (123), that $\mu = {}^{*}\mu_n$, which proves the last of the assertions of (138).

Remark. Suppose that the infinite convolution ${}^{*}\mu_n$ is convergent, i.e., that the series $\sum x_n(t)$ is convergent in ϕ-measure. Then, as mentioned in §35, the sequence $x^1(t), x^2(t), \ldots$ of the partial sums contains a subsequence which converges at every point t not contained in a certain set of ϕ-measure 0. Actually, it can be shown that, due to the particular structure (133), (134) of the terms of $\sum x_n(t)$, it is unnecessary to select a subsequence, i.e., that convergence in ϕ-measure implies convergence almost everywhere in the present case. This fact, which is a straightforward consequence of the definition of a Lebesgue integral with reference to an infinite product measure, will not be proved here, since it is irrelevant for the theory of the distributions μ on X or of their transforms (10).

39. As an application of (138), it will now be proved that, as stated at the beginning of §33,

$$\sum \{\mu_n\} < \infty \quad \text{if } {}^{*}\mu_n \text{ is convergent.} \tag{139}$$

In terms of the functions $x_1(t), x_2(t), \ldots$ and of any fixed constant c, define the functions $u_1(t), u_2(t), \ldots$ by placing

$$u_n(t) = \begin{cases} x_n(t) & \text{if } |x_n(t)| \leq 1, \\ c & \text{if } |x_n(t)| > 1. \end{cases} \tag{140}$$

Then it is clear that (134) remains valid if x_n is replaced by u_n. This means that the function $u_n(t)$ of $t = t_1.t_2. \dots$ can be thought of as a function, $u_n(t_n)$, of the n-th coordinate, t_n, of the point t of $T = T_1.T_2. \dots$; so that, corresponding to (133),

(141) $\qquad u_n(t) = u_n(t_n)$, where $t = t_1.t_2. \dots$.

Hence it is easy to show that

(142) if $\sum x_n(t)$ is convergent in ϕ-measure, then $\sum u_n(t)$ is,

where ϕ denotes the product measure (129) on (127).

In fact, suppose that $\sum x_n(t)$ is convergent (on T) in ϕ-measure. This means, by § 35, that there exists for every $q > 0$ and for every $p > 0$ an $N = N_q^p$ such that the ϕ-measure of the set of those points t at which $|x_n(t) + \dots + x_m(t)|$ exceeds p has a ϕ-measure less than q whenever $m > n > N$. In particular, the set of those points t at which $|x_j(t)|$ exceeds p has a ϕ-measure less than q whenever $j > N$. This, when combined with (140), shows that the set of those points t at which $u_j(t) = x_j(t)$ is violated by at least one $j > N$ has a ϕ-measure less than q, provided that p is chosen sufficiently small ($p < 1$). It follows therefore from the condition defining $N = N_q^p$, that the set of those points t at which $|u_n(t) + \dots + u_m(t)|$ exceeds p has a ϕ-measure less than 2q whenever $m > n > N$, provided that p is sufficiently small. This, when compared with §35, completes the proof of (142).

According to (133) and (141), both x_n and u_n are functions of the position t_n on the space T_n on which the measure is $\phi_n = \phi_n(F_n)$. Hence, if $\mu_n = \mu_n(x)$ and $\lambda_n = \lambda_n(x)$ respectively denote the ϕ_n-distributions of $x_n(t_n)$

and $u_n(t_n)$, it is seen from (140), by applying the definition (114) once to $\mu = \mu_n$, $\phi = \phi_n$, $x(t) = x_n$ and once to $\mu = \lambda_n$, $\phi = \phi_n$, $x(t) = u_n$, that $\lambda_n(x)$ is the distribution for which

$$d\lambda_n(x) = \begin{cases} 0 \text{ if } |x| > 1 \text{ and } x \neq c, \\ d\mu_n(x) \text{ if } |x| \leqq 1 \text{ and } x \neq c, \\ \{\mu_n\} + d\mu_n(x) \text{ if } x = c, \end{cases}$$

where $\{\mu_n\}$ denotes the non-negative value defined by (90). In particular, the spectrum $S(\lambda_n)$ is situated between the two points $x = \pm(1 + |c|)$ for every n. It follows therefore from (104) that a necessary condition for the convergence of $*\lambda_n$ consists in the convergence of the series $\sum (\lambda_n)_1$, where $(\lambda_n)_1 = \int x \, d\lambda_n(x)$, by (30); so that, according to the preceding representation of $d\lambda_n(x)$,

$$(\lambda_n)_1 = (c \{\mu_n\} + \int_{-1}^{1} x d\mu_n).$$

Suppose now that, as assumed in (139), the convolution $*\mu_n$ is convergent. Then, if (138) is applied to both series $\sum x_n(t)$, $\sum u_n(t)$, it follows from (142) that the convolution $*\lambda_n$ is convergent. Hence, the series $\sum (\lambda_n)_1$ must converge. Consequently, if the arbitrary constant c occurring in the last formula line is chosen to be $c = \pm 1$, it follows by subtraction of the resulting two series $\sum (\lambda_n)_1$, that $\sum 2 \{\mu_n\}$, being the difference of these convergent series, is convergent. Since $\{\mu_n\} \geqq 0$, this is identical with the assertion of (139).

40. This proof of (139), or any other application of (138) to infinite convolutions, assumes that, if a sequence of distributions $\mu_1, \mu_2, \ldots$ on X is given, it is possible to find spaces $T_1, T_2, \ldots$ on which there exist functions $x_1(t_1)$, $x_2(t_2), \ldots$ and measures $\phi_1(F_1)$, $\phi_2(F_2), \ldots$ such that (117) is satisfied and μ_n becomes the ϕ_n-distribution of $x_n(t_n)$. But this situation can be realized for any given sequence $\mu_1, \mu_2, \ldots$ by placing

$$(143) \qquad x_n(t_n) = x_n, \qquad t_n = x_n, \qquad \phi_n = \mu_n.$$

In other words, the sequence of the spaces $T_1, T_2, \ldots$ is chosen to be identical with a sequence $X_1 : -\infty < x_1 < \infty$; $X_2 : -\infty < x_2 < \infty$; ... of copies of the line $X : -\infty < x < \infty$, and the given distribution μ_n on the line X is chosen to be the measure ϕ_n on $T_n = X_n$ (cf. (1), (2), (3) and the assumptions made at the beginning of § 34). It is understood that the field of the ϕ_n-measurable sets F_n on the space T_n becomes the field of the Borel sets E_n on the line X_n. Thus the infinite product measure (129) is defined by the condition

$$(144) \qquad \phi(E_1.E_2. \ldots) = \phi_1(E_1)\, \phi_2(E_2) \ldots, \text{ where } \phi_n = \mu_n,$$

and will, in accordance with (91), be denoted by $\phi = \prod \mu_n$. It applies to sets contained in a field generated by the sets $E_1.E_2. \ldots$ of the infinite product space $X_1.X_2. \ldots$. The latter is the Cartesian space with an infinite sequence of coordinates, one point of this space being

$$(145) \qquad x_1.x_2. \ldots = (x_1, x_2, \ldots), \text{ where } -\infty < x_n < \infty .$$

It is clear from the definition (114) of a ϕ-distribution, that (138) can, in the present case, be formulated as follows:

(146) An infinite convolution $*\mu_n$ is convergent if and only if the series $\sum x_n$ is convergent in $\prod \mu_n$-measure on the infinite product space $X_1.X_2. \dots$ of the points (145), where $\prod \mu_n$ denotes the infinite product measure defined by (144). If $*\mu_n$ is convergent, then the distribution $\mu = \mu_E$ or $\mu = \mu(x)$, which is represented by $*\mu_n$ on the line X: $-\infty < x < \infty$, can be obtained as follows: The function $\sum x_n$, to which the series $\sum x_n$ converges on $X_1.X_2. \dots$ in $\prod \mu_n$-measure, attains values $x = \sum x_n$ contained in E on an $(x_1, x_2, \dots)$-set having the $\prod \mu_n$-measure μ_E, where E is any Borel set on the line X. (It is understood that the function $\sum x_n$ remains undetermined on an $(x_1, x_2, \dots)$-set of $\prod \mu_n$-measure 0).

The assertions of (146) can in various ways be combined with Borel's celebrated alternative of "either everything or nothing"; in fact, when such a combination is possible, (146) supplies a corresponding alternative with regard to properties of infinite convolutions $*\mu_n$. An instance of results of this type will be given in §41.

For the applications in question, Borel's disjunctive principle can be formulated as follows:

(147) Let $\prod \mu_n$ denote the infinite product measure on the infinite product space $X_1.X_2. \dots$ of the points (145), where $\mu_1, \mu_2, \dots$ is any sequence of distributions on a line X (the infinite convolution $*\mu_n$ need not be convergent).

Let Q and R be two complementary "qualities" such that, except for an $(x_1, x_2, \ldots)$-set of $\prod \mu_n$-measure 0, every point $(x_1, x_2, \ldots)$ of $X_1 . X_2 . \ldots$ possesses exactly one of the two properties Q,R. Suppose further that, if $(x_1', x_2', \ldots)$ and $(x_1'', x_2'', \ldots)$ are two points not on the excluded set of $\prod \mu_n$-measure 0, and if there exists to these two points an N such that $x_n' = x_n''$ for every $n > N$, then either Q or R is assigned to both points. Finally, let (Q) and (R) respectively denote the set of those points $(x_1, x_2, \ldots)$ of $X_1 . X_2 . \ldots$ to which Q and R are assigned. Then (Q), hence (R), must possess either the $\prod \mu_n$-measure 0 or the $\prod \mu_n$-measure of the whole $(x_1, x_2, \ldots)$-space, i.e., the $\prod \mu_n$-measure 1.

The proof of this alternative depends on a detailed inspection of the constructive definition of an infinite product measure; it will not be given here. Incidentally, the fact quoted in the Remark at the end of §38 can readily be obtained from this alternative.

41. As an application of (146) and (147) together, the following fact will now be proved:

(148) If every distribution μ_n occurring in a convergent convolution $*\mu_n$ is purely discontinuous, then the distribution $*\mu_n$ is pure (it can be any of the three pure types defined in §4).

First, the assumption that every μ_n is purely discontinuous means, by §4, that, if $p_n^1, p_n^2, \ldots$ is the (infinite or finite, but certainly not vacuous) sequence of the points x contained in the point spectrum, $P(\mu_n)$, of μ_n, and if $\phi_n = \phi_n(E)$ denotes the set-function $\mu = \mu_E$ belonging to the point-function $\mu = \mu(x)$ when $\mu = \mu_n$, then $\phi_n(p_n^1) + \phi_n(p_n^2) + \ldots = 1$ holds for every n (the points $x = p_n^1, p_n^2, \ldots$ are here thought of as Borel sets E on the line X : $-\infty < x < \infty$). Since, according to (143) and (138), the ϕ_n-distribution of x_n is μ_n, this implies that the set of those points x of the line X which are distinct from all the points $x = p_n^1, p_n^2, \ldots$ is of ϕ_n-measure 0 for every fixed n.

Let M denote the least modul containing every point of every $P(\mu_n)$. This means that M is the least (at most enumerable) set containing the value $x = 0$ and every value $x = p_n^k$, and having the property that it contains the sum and the difference of any two values contained in it. Thus it is clear that, if M(+)E denotes, as in §24, the vectorial sum, then M(+)E is at most enumerable whenever E is, and that M(+)E is of the Euclidean measure 0 whenever E is (by "Euclidean measure" is meant the ordinary Lebesgue measure on the x-line).

Suppose now that $*\mu_n$ is convergent. Then, according to (146), the series $\sum x_n$ converges on $X_1.X_2. \ldots$ in $\prod \phi_n$-measure, where $\phi_n = \mu_n$, and the function $\sum x_n$ of (145), thus defined on $X_1.X_2. \ldots$, has the $\prod \phi_n$-distribution $*\mu_n$. On the other hand, the purely discontinuous distribution $\mu_n = \phi_n$ on X is the ϕ_n-distribution of the function x_n of the position x_n on X_n. Hence, it is clear from the definition of the modul M, that the set M(+)E belonging to an arbitrarily fixed x-set E has the following property:

If the function $\sum x_n$ attains at two points (145) of the infinite product $X_1 . X_2 . \ldots$, say at its points $(x_1', x_2', \ldots)$ and $(x_1'', x_2'', \ldots)$, the values x' and x'' respectively, and if the two points are such that there exists an N for which $x_n' = x_n''$ holds whenever $n > N$, then the value $x = x'$ cannot be contained in the x-set $M(+)E$ unless the value $x = x''$ also is contained in $M(+)E$. This makes applicable the alternative expressed by the assertion of (147).

In fact, as pointed out before, $M(+)E$ is at most enumerable whenever E is, and $M(+)E$ is of Euclidean measure 0 whenever E is. Hence, $*\mu_n$ being the $\prod \mu_n$-distribution of the function $\sum x_n$ on $X_1 . X_2 . \ldots$, it is clear from (147) and from the definition of a ϕ-distribution (cf. §34), that, if E is either at most enumerable or of Euclidean measure 0, then the distribution $\mu = \mu_E$ represented by $\mu = *\mu_n$ attains for the set $G = M(+)E$ a value μ_G which must be either 0 or 1. Furthermore, since the point $x = 0$ is contained in M, it is clear from the definition of a vectorial sum (§ 24), that every E is a subset of the corresponding $G = M(+)E$. Thus it is clear that there are only the following possibilities left: The distribution $\mu = \mu_E$, where $\mu = *\mu_n$, attains

(i) the value $\mu_E = 1$ for a suitable enumerable set E;

(ii) the value 0 for every enumerable set E but the value $\mu_E = 1$ for a suitable Borel set E of Euclidean measure 0;

(iii) the value 0 for every Borel set E of Euclidean measure 0.

Since (i), (ii) and (iii) are precisely the properties which, according to §4, define a distribution μ to be purely discontinuous, purely singular and absolutely continuous respectively, it follows that $\mu = *\mu_n$ is pure. The remaining assertion of (148), namely that all three cases can actually occur, is shown by simple examples; cf., for instance, § 62 below.

Semi-groups

42. If μ^n denotes, as in (123), the n-th partial convolution of an infinite convolution $*\mu_n$, then $\mu^{n+1} = \mu^n * \mu_{n+1}$, by (52). In what follows, this process will be extended to the case in which the index n is replaced by a continuous parameter. The distribution sequenses (123) relate to the resulting continua of distributions in the same way as the multiplication of numbers relates to the multiplication of elements in a one-parameter semi-group.

As at the beginning of §2, let λ be any non-decreasing bounded function, normalized by

(149) $\lambda(-\infty) = 0$ and $\lambda(x - 0) = \lambda(x)$; $\lambda(\infty) < \infty$

(it need not be a distribution, since (149) allows that the second of the three conditions (8) be violated by $\mu = \lambda$). It will be shown that, whether λ is or is not a distribution, there exist to λ two distributions, say μ_I and μ_{II}, such that, if μ_I and μ_{II} are denoted by exp λ and Exp λ, the transforms $\mu_I{}^0(y) = (\exp \lambda)^0(y)$ and $\mu_{II}{}^0(y) = (\mathrm{Exp})^0(y)$ have logarithms given by

(150) $$\log (\exp \lambda)^0(y) = \int (e^{ixy} - 1)\, d\lambda(x)$$

and

(151) $$\log (\mathrm{Exp}\, \lambda)^0(y) = \int x^{-2}(e^{ixy} - 1 - ixy)\, d\lambda(x)$$

respectively. It is understood that, corresponding to the relation $z^{-2}(e^z - 1 - z) \to -\frac{1}{2}$, where $z \to 0$, the integrand of (151) is meant to be

(152) $$x^{-2}(e^{ixy} - 1 - ixy) = -\tfrac{1}{2}y^2 \qquad \text{if } x = 0$$

(whether $y \neq 0$ or $y = 0$); so that the integrand of (151) is a continuous function of x and y together. Needless to say, exp and Exp denote operators, turning the function $\lambda = \lambda(x)$ into certain distributions, $\mu_I = \mu_I(x)$ and $\mu_{II} = \mu_{II}(x)$. On the other hand, log w in (150) and (151) denotes, of course, a logarithm of the complex number w, and the branch of the logarithm is assigned by the single-valued expressions on the right of (150) and (151).

43. For instance, if $\lambda = \varepsilon$, the assertions are that there exist two distributions μ, say $\mu = \varkappa = \varkappa(x)$ and $\mu = \theta = \theta(x)$, for which the transform (10) becomes

$$(153) \qquad \varkappa^{o}(y) = e^{e^{iy} - 1}, \qquad \text{where } \varkappa = \exp \varepsilon,$$

and

$$(154) \qquad \theta^{o}(y) = e^{-\frac{1}{2}y^2}, \qquad \text{where } \theta = \operatorname{Exp} \varepsilon,$$

respectively. In fact, if λ is the unit distribution (15), then, while the integral (150) reduces to $e^{iy} - 1$, the integral (151) becomes $-\frac{1}{2}y^2$ in virtue of (152). Actually, the distributions to be defined by (153) and (154), the "Poissonian" distribution $\mu = \varkappa$ and the "Gaussian" distribution $\mu = \theta$, turn out to be such as to lead, by a process of stratification, to the distributions exp λ, Exp λ belonging to an arbitrary $\lambda = \lambda(x)$. In fact, it will be shown that exp λ and Exp λ result from $\varkappa$ and θ for every λ, if an averaging process (to be carried out with reference to an arbitrary weight-function, depending on the arbitrary λ) is combined with the two trivial processes defined by (17) and (16); and so, in particular, with the convolution process belonging to the translated unit distribution

(155) $\quad {}^{c}\varepsilon(x) = \varepsilon(x - c)$, where $\log {}^{c}\varepsilon^{\circ}(y) = icy$.

The existence of the two fundamental distributions, $\varkappa = \exp \varepsilon$ and $\theta = \operatorname{Exp} \varepsilon$, can be proved directly, as follows: If n runs through all non-negative integers, the sum of the reciprocal values of the positive numbers $n!e$ is 1. Hence, there exists a distribution $\mu = \mu(x)$ which has at $x = n$ the jump $1/(n!e)$ and is constant for $-\infty < x < 0$ and for $n < x < n + 1$, where $n = 0,1,\ldots$. According to (10), the transform $\mu^{\circ}(y)$ of this purely discontinuous distribution is the sum of all terms $e^{iny}/(n!e)$. Since this infinite series has the sum e^{w}/e, where $w = e^{iy}$, it follows that (153) is satisfied by the distribution just defined. This proves the existence of the distribution $\varkappa$. It also follows that, the function (153) of y being regular-analytic at $y = 0$, the distribution $\varkappa$ has finite moments $(\varkappa)_k$ of arbitrarily high order (cf. the beginning of § 16). In particular, from (32) and (153),

(156) $\quad (\varkappa)_1 = 1, \quad (\varkappa)_2 = 2; \qquad$ hence $[\varkappa] = 1$,

by (40). Moreover, $\varkappa^{\bullet}(x) = \varkappa(x - 1)$, by (38). According to (16), this means that $\varkappa^{\bullet\circ}(y) = e^{iy}\varkappa^{\circ}(y)$. It follows therefore from (153) that $\varkappa^{\bullet\circ}(y) = 1 - \frac{1}{2}y^2 + o(y^2)$, hence $\log \varkappa^{\bullet\circ}(y) = -\frac{1}{2}y^2 + o(y^2)$, as $y \longrightarrow 0$. This obviously implies that, as $n \longrightarrow \infty$, the function $n \log \varkappa^{\bullet\circ}(n^{-\frac{1}{2}}y)$ of y tends to the function $-\frac{1}{2}y^2$ uniformly on every fixed bounded y-interval, which means that, on every such interval, the n-th power of the function $\varkappa^{\bullet\circ}(n^{-\frac{1}{2}}y)$ of y tends to the exponential (154). It follows therefore from the continuity theorem, formulated at the end of §10, that, in order to prove the existence of a distribution θ satisfying (154), it is sufficient to ascertain that the n-th power of the function $\varkappa^{\cdot 0}(n^{-\frac{1}{2}}y)$ of y is the transform of a distribution for every n. But this is clear from (82) and (17), since the function $\varkappa^{\cdot 0}(y)$ of y is the transform of a distribution, $\varkappa^{\cdot}$. Thus the distribution θ appears reduced to the distribution $\varkappa$. (In this regard, it is irrelevant that θ will, in § 54 , be given explicitly).

44. If $\lambda(x)$ = const., then, on the one hand, const. = 0 in view of (149) and, on the other hand, both integrals (150), (151) vanish for every y. It follows therefore from (15) that, if $\lambda(x)$ = const., the distributions assigned by (150) and (151) are

$$(157) \qquad \exp 0 = \varepsilon, \qquad\qquad \operatorname{Exp} 0 = \varepsilon.$$

If the existence of the distributions assigned by (150) and (151) has been established for the non-decreasing bounded functions $\lambda = \lambda_1$ and $\lambda = \lambda_2$, then the existence of the distributions assigned for the non-decreasing bounded function $\lambda = \lambda_1 + \lambda_2$ follows from

$$(158) \quad \exp(\lambda_1 + \lambda_2) = (\exp \lambda_1)*(\exp \lambda_2), \quad \operatorname{Exp}(\lambda_1 + \lambda_2) = (\operatorname{Exp} \lambda_1)*(\operatorname{Exp} \lambda_2).$$

This is clear from the product rule (53), since, on the one hand, both assignments (150), (151) are logarithmic and, on the other hand, both are such as to be additive in λ.

If a non-decreasing bounded function has a jump at $x = 0$, then it obviously has the form $p\varepsilon + \lambda$, where λ is a non-decreasing bounded function having no jump at $x = 0$, and p is a positive constant, representing the jump given at $x = 0$. Since (154) and (17) imply the existence of the distribution Exp $(p\varepsilon)$, it follows from (158) that, in the proof for the existence of the distribution Exp λ for an arbitrary λ, it will be sufficient to consider the case in which λ is continuous at $x = 0$. It will turn out that in this case the existence of Exp λ can be reduced to the existence of exp λ for an arbitrary λ. (Incidentally, the existence of Exp λ in the complementary case, where $\lambda = p\varepsilon$, is reduced to the existence of exp λ for $\lambda = \varepsilon$, i.e., to the existence of $\varkappa$; cf. the end of §44).

If $\lambda = \lambda_1$ and $\lambda = \lambda_2$ are monotone functions satisfying (149), and if $\lambda_1 * \lambda_2(x)$ is defined by placing $\mu_1 = \lambda_1$, $\mu_2 = \lambda_2$ in (48), it is clear that $\lambda = \lambda_1 * \lambda_2$ is a monotone function satisfying (149). It is also clear that (53) remains valid for $\mu_1 = \lambda_1$, $\mu_2 = \lambda_2$, if $\lambda^{o}(y)$ is defined by placing $\mu = \lambda$ in (10). In particular, $\lambda_1 * \lambda_2(\infty) = \lambda_1(\infty)\,\lambda_2(\infty)$, since $\lambda^{o}(0) = \lambda(\infty)$. And $\lambda(\infty)$ is the total variation of any λ, since $\lambda(-\infty) = 0$. If λ is any monotone function satisfying (149), let $\lambda^{1}, \lambda^{2}, \ldots$ be defined by placing $\lambda^{n} = \lambda^{n-1} * \lambda$ and $\lambda^{1} = \lambda$. Then it is seen by complete induction that (149) is satisfied by $\lambda = \lambda^{n}$, and that the total variation of λ^{n} is the n-th power of $\lambda(\infty) = \lambda^{o}(0)$, since $(\lambda^{n})^{o}(y)$ is the n-th power of $\lambda^{o}(y)$. Hence, if the trivial case $\lambda(\infty) = 0$, for which exp λ and Exp λ are taken care of by (157), is excluded, then $\lambda^{n}(x)/\lambda^{n}(\infty)$ is a distribution and has the transform represented by the n-th power of $\lambda^{o}(y)/\lambda^{o}(0)$, where $n = 1,2,\ldots$

According to (15), all of this remains valid for $n = 0$, if λ^0 is defined to be ε for every λ. Since the sum of the positive constants $(n!e)^{-1}$, where $n = 0,1,2,\ldots$, is 1, it follows that the first of the series

$$\sum_{n=0}^{\infty} \frac{1}{n!e}\,\frac{\lambda^{n}(x)}{\lambda^{n}(\infty)}, \qquad \sum_{n=0}^{\infty} \frac{1}{n!e}\left(\frac{\lambda^{0}(y)}{\lambda^{0}(0)}\right)^{n}$$

is a distribution, and that the second series is the transform of this distribution. But the sum of the second series is e^{w}/e, where $w = \lambda^{0}(y)/\lambda^{0}(0)$. Hence, if λ^0 is substituted from the case $\mu = \lambda$ of (10), it follows that the distribution represented by the first series has a transform possessing a logarithm identical with the integral (150). This proves that the assignment (150) defines a distribution exp λ for every monotone λ satisfying (149).

Next, if $h(u)$ is an abbreviation for $u^{-2}(e^{iu} - 1 - iu)$, then $h(u) = -\frac{1}{2} + o(1)$ as $u \to \pm 0$ and $h(u) = O(|u|^{-1})$ as $u \to \pm\infty$. This obviously implies that, as $n \to \infty$,

$$\int\limits_{1/n<|x|<\infty} x^{-2}(e^{ixy} - 1 - ixy)\,d\lambda \longrightarrow \int\limits_{0<|x|<\infty} x^{-2}(e^{ixy} - 1 - ixy)\,d\lambda$$

holds uniformly on every fixed bounded y-interval, if $\lambda = \lambda(x)$ is any monotone function satisfying (149). But the integral on the right is identical with the integral (151), if λ has no jump at $x = 0$, i.e., if (152) does not contribute to (151). On the other hand, if $\lambda_n(x)$ denotes a non-decreasing bounded function for which $d\lambda_n(x) = 0$ or $d\lambda_n(x) = d\lambda(x)$ according as $|x| < 1/n$ or $|x| > 1/n$, then the integral on the left of the last formula line can be thought of as a complete integral (151). It follows therefore from the continuity theorem, formulated at the end of § 10, that, in order to prove the existence of a distribution Exp λ satisfying (151), it is sufficient to show the existence of the distributions Exp λ_n, provided that λ is assumed to have no jump at $x = 0$. But it was seen after (158) that this assumption can be granted. Consequently, the existence of the distribution Exp λ will be proved for every λ, if it is proved for every λ which is constant on some open interval containing the point $x = 0$. However, if there exists such an interval, then there exists a non-decreasing bounded function, say λ_0, such that $d\lambda_0(x) = x^{-2} d\lambda(x)$ for $-\infty < x < \infty$. Hence, the integral (151) can be written in the form

$$\int (e^{ixy} - 1)\, d\lambda_0(x) + icy, \text{ where } c = - \int x^{-1}\, d\lambda(x)$$

(the integral defining the constant c is convergent, since λ is constant near $x = 0$). It follows therefore from (150), (155) and (53), that the assignment (151) is satisfied by Exp $\lambda = {}^{c}\varepsilon * \exp \lambda_0$; that is, by a distribution which, according to (55), differs from the distribution exp λ_0 only in a translation.

45. This proves that (150), (151) define distributions exp λ, Exp λ for any monotone λ satisfying (149). It will now be shown that any semi-group of the type indicated at the beginning of §42 is expressible in terms of these distributions and of (155); distributions which, in turn, generate semi-groups for arbitrary choices of λ and c in (150), (151) and (155) respectively.

Let there be given for every value of a positive parameter t a distribution $\mu^t = \mu^t(x)$ on the line $-\infty < x < \infty$ in such a way that

$$(160) \qquad \mu^{u+v} = \mu^{u} * \mu^{v}, \qquad (0 < u < \infty,\ 0 < v < \infty)$$

and

$$(160') \qquad \mu^{t} \rightarrow \varepsilon \text{ as } t \longrightarrow +0; \text{ cf. } (15).$$

Then the sheaf $0 < t < \infty$ of distributions μ^{t} will be called a semi-group, and $\mu^{1} = (\mu^{t})_{t=1}$ its generator. The latter will simply be denoted by $\mu = \mu(x)$. It turns out that the correspondence between these sheets of distributions μ^{t} and their generators μ^{1} is one-to-one. Let therefore a given distribution μ be called a generator if there exists a semi-group μ^{t} satisfying $\mu^{1} = \mu$. The object of the following considerations is the determination of all those distributions μ which are generators.

First, it is clear that, if $\lambda(x)$ is a monotone function having the properties (149), then $t\lambda(x)$ is a monotone function having the same properties, where t denotes any positive number. Hence, exp $(t\lambda)$ and Exp $(t\lambda)$ represent distributions for every λ and for every t. According to (158), both $\mu^{t} = \exp\ (t\lambda)$ and $\mu^{t} = \mathrm{Exp}\ (t\lambda)$ satisfy (160) for every fixed λ. Moreover, it is clear from (150), (151), (28) and (56), that (160') is satisfied in both cases. Similarly, it is obvious from (155) and (87) that (160) and (160') are satisfied by $\mu^{t} = {}^{ct}\varepsilon$, if c is arbitrarily fixed. Finally, it is seen from (56) and (52) that, if either of the requirements (160), (160') is satisfied by two sheaves of distributions, say by $\mu^{t} = \mu_{1}^{t}$ and by $\mu^{t} = \mu_{2}^{t}$, then it is satisfied by the sheaf $\mu^{t} = \mu_{1}^{t} * \mu_{2}^{t}$ also. Consequently, if λ' and λ'' are arbitrary monotone functions satisfying (149), and if c is any real constant, then

$$(161) \qquad \mu = {}^{c}\varepsilon * (\exp \lambda') * (\mathrm{Exp}\ \lambda'')$$

is a distribution representing the generator, μ^{1}, of a semi-group μ^{t};

$0 < t < \infty$. And what will be shown is the converse, namely the fact that a given distribution, μ, is representable in the form (161) whenever it is a generator, μ^1.

46. If r is any fixed positive number, then (150) and (151) show that the contributions of the ranges $|x| \leqq r$ and $|x| > r$ to $\log(\exp\lambda')^0(y)$ and $\log(\operatorname{Exp}\lambda'')^0(y)$ are

$$\int_{|x| \leqq r} (e^{ixy} - 1)\, d\lambda' \text{ and } \int_{|x| > r} x^{-2}(e^{ixy} - 1 - ixy)\, d\lambda''$$

respectively. Consequently, these contributions can be written as

$$iay + \int_{|x| \leqq r} x^{-2}(e^{ixy} - 1 - ixy)\, d\lambda'_0 \text{ and } iby + \int_{|x| > r} (e^{ixy} - 1)\, d\lambda''_0$$

respectively, where a and b are real constants and λ'_0, λ''_0 denote functions of the same type as the functions λ', λ'', that is, non-decreasing bounded founctions (cf. the last formula line of § 44). Thus it follows by the reasoning applied at the end of § 44, that the contributions of the complementary ranges, $|x| \leqq r$ and $|x| > r$, to $\exp\lambda'$ and $\operatorname{Exp}\lambda''$ can respectively be represented as $\overset{a}{\varepsilon} * \operatorname{Exp}\lambda'_0$ and $\overset{b}{\varepsilon} * \exp\lambda''_0$. It follows therefore from (52), (87) and (158) that, if a distribution μ can be represented in the form (161), then the number c and the non-decreasing bounded functions λ', λ'' occurring in (161) can be so chosen that λ' and λ'' are constant on the interval $|x| < 1$ and on both of the complementary half-lines, $|x| > 1$, respectively. Thus it is clear from (150), (151) and (155), that the representation (161) is equivalent to

$$(162)\quad \log\ \mu^{o}(y) = icy + \left(\int - \int_{-1}^{1}\right)(e^{ixy} - 1)\, d\lambda + \int_{-1}^{1} x^{-2}(e^{ixy} - 1 - ixy)\, d\lambda,$$

where c is a real constant and λ denotes a monotone function satisfying (149).

Accordingly, what has been verified above is that, if there exists a pair (c,λ) satisfying (162) for a given distribution μ, then μ is a generator. And the converse to be proved is equivalent to the statement that there exists for every given generator μ a pair (c,λ) satisfying (162).

Incidentally, the replacement of (161) by the normalized representation (162) is actually such as to make the correspondence between generators μ and the pairs (c,λ) a one-to-one correspondence. In other words, if (c_1,λ_1) and (c_2,λ_2) belong to the same μ, then $c_1 = c_2$ and $\lambda_1 = \lambda_2$.

In order to prove this, let $a = c_1 - c_2$ and $\alpha = \lambda_1 - \lambda_2$. It then follows from (162) by subtraction, that

$$(163)\quad 0 = iay + \left(\int - \int_{-1}^{1}\right)(e^{ixy} - 1)\, d\alpha + \int_{-1}^{1} x^{-2}(e^{ixy} - 1 - ixy)\, d\alpha.$$

Since λ_1 and λ_2 are monotone functions satisfying (149), the function $\alpha = \alpha(x)$ is of bounded variation, is continuous from the left, and vanishes as $x \longrightarrow -\infty$. Accordingly, the assertion is that (163), where $\int |d\alpha| < \infty$, cannot hold (for every y) unless $a = 0$ and $\alpha(x - 0) = \text{const}$.

For a fixed real number b, integrate the identity (163) between the limits $y = b - 1$ and $y = b + 1$ and divide by 2. Then, if b is renamed y, the result appears in the form

$$0 = iay + \left(\int - \int_{-1}^{1}\right) e^{ixy}(\sin x/x - 1)\, d\alpha + \int_{-1}^{1} x^{-2}(e^{ixy} \sin x/x - 1 - ixy)\, d\alpha,$$

where sin x/x = (sin x)/x. If this is subtracted from (163), it follows that

$$0 = \left(\int - \int_{-1}^{1}\right) e^{ixy}(1 - \sin x/x)\, d\alpha + \int_{-1}^{1} x^{-2} e^{ixy}(1 - \sin x/x)\, d\alpha.$$

This can be written in the form $0 = \int e^{ixy}\, d\beta$, where

$$d\beta(x) = \begin{cases} (1 - \sin x/x)d\alpha(x) \text{ if } |x| > 1, \\ x^{-2}(1 - \sin x/x)\, d\alpha(x) \text{ if } |x| \leqq 1. \end{cases}$$

Since $(1 - \sin x/x)$ and $x^2(1 - \sin x/x)$ are regular at $x = 0$, and since $\int |d\alpha| < \infty$, it is clear that $\int |d\beta| < \infty$. It follows therefore from (19*) that $\beta(x - 0) = \text{const.}$, since $0 = \int e^{ixy}\, d\beta$ for every y. But $\beta(x - 0) =$ const. and the preceding formula line obviously imply that $\alpha(x - 0) =$ const. And $a = 0$ now follows from (163).

46'. In the parallelism referred to at the beginning of §42, the relation (97) corresponds to (160'). However, while it is possible to prove that (97) must hold if $*\mu_n$ is assumed to be convergent, it is impossible to disprove the negation of (160') if (160) is supposed to be satisfied. By this is meant that an admission of the continuum hypothesis leads to sheaves μ^t which satisfy (160) and violate (160'). In fact, while every measurable, and every locally bounded (or half bounded) solution of the functional equation

$$f(u + v) = f(u) + f(v) \tag{164}$$

is known to be represented by the linear functions $f(t) = f(1)t$, the continuum hypothesis supplies non-measurable solutions $f(t)$ of (164). But if $f(t)$ is any real-valued solution of (164), it is clear from (87) that (160) is satisfied by $\mu^t = {}^{f(t)}\varepsilon$; while (160') will hold only if $f(t)$ is so chosen as to satisfy $f(t) = o(1) = O(1)$ as $t \longrightarrow 0$. This is the reason that both (160) and (160') have been required in the definition of a semi-group μ^t.

According to (53) and (160),

$$(165) \qquad (\mu^{u+v})^o(y) = (\mu^u)^o(y) \cdot (\mu^v)^o(y).$$

This implies that $(\mu^{nt})^o(y)$ is the n-th power of $(\mu^t)^o(y)$ for $n = 1,2,\ldots$ and for every t. On the other hand, it is clear from (160') and from the uniformity result of §10, that there exists for every $p > 0$ a sufficiently small $q_p > 0$ such that $|(\mu^t)^o(y)| > \frac{1}{2}$ whenever $-p < y < p$ and $0 < t < q_p$. Consequently,

$$(166) \quad |(\mu^t)^o(y)| > (\tfrac{1}{2})^n \text{ if } -p < y < p \text{ and } 0 < t < nq_p.$$

Since p can be chosen arbitrarily large, it follows, by letting $n \longrightarrow \infty$ when p is fixed, that

$$(167) \qquad (\mu^t)^o(y) \neq 0 \text{ for } -\infty < y < \infty \text{ and } 0 < t < \infty .$$

According to §10, every transform is a continuous function of y. Hence, if a transform is distinct from 0 for every y, there exists for $-\infty < y < \infty$ exactly one continuous function representing a logarithm of the transform for every y and attaining the value 0 for $y = 0$; cf. (12). This, when compared with (167), defines

$$(168) \qquad f_y(t) = \log (\mu^t)^o(y), \qquad (f_0(t) = 0),$$

as a unique continuous function of y for $-\infty < y < \infty$, if t is fixed. On the other hand, (165) means that the function (168) satisfies (164), if y is fixed. But (166) and (11) imply that the solution (168) of (164) is bounded on every bounded t-interval, if y is fixed. Consequently, (168) is a linear function of t for every fixed y; so that $f_y(t) = f_y(1)t$. If this is substituted into (168), it follows from (165) that

$$(169) \qquad (\mu^t)^o(y) = (\mu^o(y))^t, \text{ where } \mu = \mu^1;$$

so that, according to (167),

$$(170) \qquad \mu^o(y) \neq 0 \text{ for } -\infty < y < \infty.$$

It is understood that the exponential, $(\ \ )^t = e^{t \log (\ \ )}$, on the right of (169) is defined in terms of the continuous logarithm specified by (168).

Accordingly, (160) and (160') together are equivalent to (169). In fact, it is clear from (28) and (15) that (160') is implied by (169). It also follows from (169) and (170) that

$$(171) \qquad \frac{(\mu^t)^o(y) - 1}{t} \longrightarrow \log \mu^o(y) \text{ as } t \longrightarrow 0$$

holds uniformly on every fixed bounded y-interval (in fact, both $w = \mu^o(y)$ and $1/w$ are bounded, and therefore $(w^t - 1)/t \longrightarrow \log w$ holds uniformly, on every such y-interval).

47. These results, proved for an arbitrary semi-group μ^t, will now be used in order to prove that every given generator $\mu = \mu^1$ is representable in the form (161).

In terms of the semi-group which is generated by a given $\mu = \mu^1$, define, for every fixed $t > 0$ and for $-\infty < x < \infty$, two monotone functions, say $\lambda'_t = \lambda'_t(x)$ and $\lambda''_t = \lambda''_t(x)$, by placing

$$(172) \qquad d\lambda'_t(x) = \begin{cases} 0 \text{ if } |x| \leqq 1, \\ t^{-1} d\mu^t(x) \text{ if } |x| > 1 \end{cases}$$

and

$$(173) \qquad d\lambda''_t(x) = \begin{cases} t^{-1} x^2 d\mu^t(x) & \text{if } |x| < 1, \\ 0 & \text{if } |x| > 1 \end{cases}$$

(where $d = d_x$). Since μ^t, where t is fixed, is a distribution, it can be assumed that (149) is satisfied by both $\lambda = \lambda'_t$ and $\lambda = \lambda''_t$.

According to (172) and (173), the logarithms, (150) and (151), of the transforms of the distributions $\exp \lambda'_t$ and $\operatorname{Exp} \lambda''_t$ are respectively represented by

$$t^{-1}\left(\int - \int_{-1}^{1}\right)(e^{ixy} - 1) d\mu^t(x) \text{ and } t^{-1} \int_{-1}^{1} (e^{ixy} - ixy - 1)\, d\mu^t(x).$$

Hence, if c_t is an abbreviation for

$$(173) \qquad c_t = t^{-1} \int_{-1}^{1} x\, d\mu^t(x),$$

the sum of these two logarithms is identical with

$$t^{-1} \int (e^{ixy} - 1)\, d\mu^t(x) - ic_t y,$$

and so, in view of (10), (12) and (155), with

$$t^{-1}(\mu^t)^\circ(y) - t^{-1} - \log({}^{c_t}\varepsilon)^\circ(y).$$

Accordingly, the sum of the logarithms of the transforms of the three distributions ${}^{c_t}\varepsilon$, $\exp \lambda'_t$, $\operatorname{Exp} \lambda''_t$ is identical with the quotient on the left of (171). Since the relation (171) holds uniformly on every fixed bounded y-interval, it follows that the same is true of the relation

$$(174)\quad ({}^{c_t}\varepsilon)^\circ(y) . (\exp\lambda'_t)^\circ(y) \;.\; (\mathrm{Exp}\,\lambda''_t)^\circ(y) \to \mu^\circ(y) \text{ as } t \to 0;$$

a relation which, according to (82) and (28), is equivalent to

$$(175)\quad ({}^{c_t}\varepsilon) * (\exp\lambda'_t) * (\mathrm{Exp}\,\lambda''_t) \to \mu \text{ as } t \to 0.$$

While (175) indicates, it does not in itself imply, the existence of a representation (161). However, the transition from (175) to (161) can now be based onthe estimates (21), (22).

48. First , (22) and (21) show that, if t and r are arbitrarily fixed,

$$\Big(\int - \int_{-r}^{r}\Big)\, d\mu^t(x) \text{ and } \int_{-r}^{r} x^2\, d\mu^t(x)$$

are respectively majorized by the maximum of $4|1 - (\mu^t)^\circ(y)|$ on the interval $-1/r \leqq y \leqq 1/r$ and by the minimum of $3\,|1 - (\mu^t)^\circ(y)|/y^2$ on the same interval. If $r > 1$ in the first case and $r = 1$ in the second case, these estimates can, by (172) and (173), be written as

$$(176)\quad \Big(\int - \int_{-r}^{r}\Big)\, d\lambda'_t(x) < t^{-1} \underset{-1/r \leqq y \leqq 1/r}{\mathrm{Max}} |1 - (\mu^t)^\circ(y)|, \qquad (r > 1)$$

and

$$(177)\quad \int = \int_{-1}^{1} x^2 d\lambda''_t(x) \leqq 3t^{-1} \underset{-1 \leqq y \leqq 1}{\mathrm{Min}} |1 - (\mu^t)^\circ(y)|/y^2$$

respectively. But (171) holds uniformly on every fixed interval $-1/r \leqq y \leqq 1/r$, and (170) shows that the function on the right of (171) is a continuous function which, according to (12), vanishes at $y = 0$. It follows therefore from (176) and (171) that the total variation of $\lambda'_t(x)$ on the exterior of the interval $-r < x < r$ tends to 0 as $t \rightarrow 0$ and $r \rightarrow \infty$. It is similarly seen from (171) and from (177), or rather from the proof of (177), that $\int x^2 d\lambda''_t(x) = O(1)$ as $t \rightarrow 0$. Actually, this O-estimate is equivalent to $\int d\lambda''_t(x) = O(1)$ as $t \rightarrow 0$, since, according to (173), the functions $\lambda = \lambda''_t$ are constant for $-\infty < x < -1$ and for $1 < x < \infty$.

Since both $\lambda = \lambda'_t$ and $\lambda = \lambda''_t$ are monotone functions satisfying (149) for every t, it follows that, as $t \rightarrow 0$, the functions λ'_t and λ''_t are uniformly bounded for $-\infty < x < \infty$. Consequently, Helly's selection theorem is applicable. It supplies the existence of a sequence $t_1, t_2, \dots$ and of two monotone functions, say λ' and λ'', such that $t_n \rightarrow 0$ as $n \rightarrow \infty$, both $\lambda = \lambda'$ and $\lambda = \lambda''$ satisfy (149) and, if λ'_n, λ''_n respectively denote the functions λ'_t, λ''_t belonging to $t = t_n$, the functions λ'_n, λ''_n tend to λ', λ'' at every continuity point of $\lambda' = \lambda'(x)$, $\lambda''(x)$ respectively.

In addition, the total variation of the monotone function λ'_n on the exterior of the interval $-r \leqq x \leqq r$ tends to 0 uniformly for all n, as $r \rightarrow \infty$. Since λ'_n tends to λ' at every continuity point of $\lambda' = \lambda'(x)$, this implies that, as $n \rightarrow \infty$, the integral (150) belonging to $\lambda = \lambda'_n$ tends to the integral (150) belonging to $\lambda = \lambda'$; so that

$$(178) \qquad (\exp \lambda'_n)^{\circ}(y) \rightarrow (\exp \lambda')^{\circ}(y).$$

On the other hand, the monotone functions λ''_n are uniformly bounded for $-\infty < x < \infty$ and are independent of x for $-\infty < x < -1$ and for $1 < x < \infty$. Since λ''_n tends to λ'' at every continuity point of $\lambda'' = \lambda''(x)$, this implies that, as $n \longrightarrow \infty$, the integral (151) belonging to $\lambda = \lambda''_n$ tends to the integral (151) belonging to $\lambda = \lambda''$; so that

$$(179) \qquad (\mathrm{Exp}\ \lambda''_n)^{\circ}(y) \longrightarrow (\mathrm{Exp}\ \lambda'')^{\circ}(y).$$

According to (150), (151) and (170), the limit functions occurring in (178), (179) and (174) are distinct from 0 for every y. It follows therefore from (178), (179) and (174) that, if $t = t_n$, the first factor on the left of (174) must tend to a limit as $t = t_n \longrightarrow 0$, i.e., as $n \longrightarrow \infty$. Since this holds for every y, it follows from (155) that there exists a constant c such that, if c_n denotes the value c_t belonging to $t = t_n$, then $c_n \longrightarrow c$, i.e.,

$$(180) \qquad {}^{c_n}\varepsilon^{\circ}(y) \longrightarrow {}^{c}\varepsilon(y).$$

In view of (56), (53) and (28), the relations (180), (178) imply that

$$({}^{c_n}\varepsilon) * (\exp \lambda'_n) * (\mathrm{Exp}\ \lambda''_n) \longrightarrow ({}^{c}\varepsilon) * (\exp \lambda') * (\mathrm{Exp}\ \lambda'').$$

If this is compared with (175), where $t = t_n$, it follows that μ can be represented in the form (161).

Incidentally, it is seen from the proof that, in the analogy mentioned at the beginning of § 42, the three members on the right of (161) are paralleled by the three series (111).

Inversions

49. In view of (19), there arises the question as to the existence of an explicit inversion formula for (10); that is, of a formula representing an arbitrary distribution $\mu = \mu(x)$ in terms of its transform $\mu^0 = \mu^0(y)$.

In order to obtain such a formula, it will first be verified that

$$(181) \qquad 1 - 2\pi^{-1} \int \mathrm{si}\ (rt - rx)\ d\mu(t) \longrightarrow \mu(x+0) + \mu(x-0) \text{ as } r \longrightarrow \infty,$$

where $\int = \int_{-\infty}^{\infty}$ and $\mathrm{si}(u) = \int_0^u v^{-1} \sin v\, dv.$

Since $\mu(\infty) = 1$, $\mu(-\infty) = 0$ and $\mathrm{si}(\pm\infty) = \pm\frac{1}{2}\pi$, a partial integration shows that

$$(182) \qquad \int \mathrm{si}(rt)\ d\mu(t) = \tfrac{1}{2}\pi - \int \mu(t) t^{-1} \sin rt\ dt.$$

Since μ is monotone and bounded, it is clear from the second mean-value theorem that, for every fixed $a > 0$, the contribution of both half-lines $-\infty < t < -a$, $a < t < \infty$ to the integral on the right of (182) tends to 0 as $r \longrightarrow \infty$. On the other hand, since $\mu(t) = \mu(\pm 0) + o(1)$ as $t \to \pm 0$, the contribution of the interval $-a \leqq t \leqq a$ is

$$\mu(+0) \int_0^a t^{-1} \sin rt\ dt + \mu(-0) \int_{-a}^0 t^{-1} \sin rt\ dt + \int_{-a}^a o(1) t^{-1} \sin rt\ dt,$$

which, since $\mathrm{si}(u) = \int_0^u v^{-1} \sin v\ dv$, can be written in the form

$$\mu(+0)\mathrm{si}(ar) - \mu(-0)\ \mathrm{si}(-ar) + \int_{-ar}^{ar} o(1)\ t^{-1} \sin t\ dt.$$

Since $|t^{-1} \sin t| \leqq 1$ and $\mathrm{si}(\pm\infty) = \pm\frac{1}{2}\pi$, it is clear that a can so be chosen as a function of r that, as $r \longrightarrow \infty$, the function a tends to 0, the product ra to ∞, finally the expression in the last formula line to $\frac{1}{2}\pi\mu(+0) + \frac{1}{2}\pi\ \mu(-0)$.

If all of this is substituted into the integral on the right of (182), the identity (182) proves the particular case $x = 0$ of (181). But the case of an arbitrary x can be reduced to the case $x = 0$ by shifting to $t = x$ the origin of the t-axis on the left of (181). Hence, (181) is true for every x.

There is a similar relation for the case where the sum on the right of (181) is replaced by the corresponding difference, i.e., by the (non-negative) jump of the distribution μ at a point x; in fact,

$$(183)\qquad \int \frac{\sin(rt - rx)}{rt - rx}\, d\mu(t) \longrightarrow \mu(x + 0) - \mu(x - 0) \text{ as } r \longrightarrow \infty .$$

In contrast to the proof of (181), which depends on the second mean-value theorem, the proof of (183) requires only the first mean-value theorem. In fact, since $\int |d\mu(t)| < \infty$ and $|\sin| < 1$, this theorem implies that

$$(184)\qquad r^{-1}\int t^{-1}\sin(rt)\, d\mu(t) = r^{-1}\int_{-a}^{a} t^{-1} \sin(rt)\, d\mu(t) + o(1) \text{ as } r \longrightarrow \infty$$

holds for every fixed $a > 0$. On the other hand, since $\mu(t) \longrightarrow \mu(\pm 0)$ a $t \longrightarrow \pm 0$ and $z^{-1} \sin z \longrightarrow 1$ as $z \longrightarrow 0$, it is clear that a can so be chosen as a function of r that the expression preceding the $o(1)$ on the right of (184) tends to $\mu(+0) - \mu(-0)$ as $r \longrightarrow \infty$. It follows therefore from (184) that (183) is true in the case $x = 0$. Finally, the case of an arbitrary x can be reduced to the case $x = 0$ by shifting to $t = x$ the origin of the t-axis on the left of (183).

50. As a corollary of (181), it will now be verified that an explicit inversion of the transformation (10) of a distribution μ into its transform μ^{o} is given by

$$(185)\qquad \frac{\mu(x+0)+\mu(x-0)}{2} = \text{const.} + \frac{1}{2\pi}\lim_{r\to\infty}\int_{-r}^{r} \frac{1-e^{-ixy}}{iy}\,\mu^{o}(y)\,dy$$

for every x (the numerical value of the constant follows by placing $x = 0$). Correspondingly, (183) will lead to an explicit representation of the (non-negative) jump $\mu(x+0) - \mu(x-0)$ in terms of the transform, namely to

$$(186)\qquad \mu(x+0)-\mu(x-0) = \lim_{r\to\infty}\frac{1}{2r}\int_{-r}^{r} e^{-ixy}\,\mu^{o}(y)\,dy.$$

Since $(1 - e^{-z})/z \longrightarrow 1$ as $z \longrightarrow 0$, the integrand on the right of (185) remains continuous at $y = 0$, and so the existence of the integral on the right of (185) is obvious, if $r < \infty$. On the other hand, the existence of the limit on the right of (185) is part of the statement. And this limit cannot in general be replaced by the Lebesgue integral, $\int$, over the line $-\infty < y < \infty$. In fact, it cannot in general be replaced by the ordinary improper integral $\int$; that is, by the limit of the contribution of the range $-p \leqq y \leqq r$ as r and p tend to ∞ independently, rather than so as to be subject, as in the "principal" integral (185), to the restriction $r = p$.

In order to prove the necessity of this restriction, it is sufficient to show that, if the contribution of the range $-r \leqq x \leqq 0$ is omitted in (185), the resulting integral over the range $0 \leqq x \leqq r$ need not have a limit as $r \longrightarrow \infty$. But, if $x = 1$ and $\mu^{o}(y) = \cos y$, then the imaginary part of the integral just mentioned is

$$-\int_0^r (1 - \cos y)y^{-1}\cos y \, dy = \text{Const.} + o(1) - \int_1^r y^{-1}\cos^2 y \, dy \rightarrow -\infty$$

as $r \rightarrow \infty$. Hence it is sufficient to ascertain the existence of a distribution $\mu(x)$ for which $\mu^o(y)$ becomes $\cos y$. But (10) shows that such a $\mu(x)$ is represented by the step-function which is 0, $\frac{1}{2}$ or 1 according as $-\infty < x < -1$, $-1 < x < 1$ or $1 < x < \infty$.

On the other hand, the limiting average on the right of (186) always exists, and has the value represented by the left of (186), even if the range $-r \leqq y \leqq r$, its length $2r$ and the limit proces $2r \rightarrow \infty$ are replaced by a range $u \leqq y \leqq v$, by its length $v - u$ and by the limit process $v - u \rightarrow \infty$ respectively (so that not only $u = 0$, $v \rightarrow \infty$, or $u \rightarrow -\infty$, $v = 0$ but, if $v - u \rightarrow \infty$, even $u \rightarrow \infty$, $v \rightarrow \infty$ or $u \rightarrow -\infty$, $v \rightarrow -\infty$ are allowed). However, this generalization of (186), which is proved in exactly the same way as (186) itself, will not be needed.

51. The symmetry restriction in (185) will be used by writing the integral (185) in the form

$$\int_{-r}^{r} = \int_0^r \frac{1}{iy} \left\{(1 - e^{-ixy})\, \mu^o(y) - (1 - e^{ixy})\, \mu^o(-y)\right\} dy.$$

Replace x in (10) by t and substitute the resulting representations of $\mu^o(y)$ and $\mu^o(-y)$ into the last formula. This gives

$$\int_{-r}^{r} = \int_{0}^{r} \frac{1}{iy}\left\{\int (e^{ity} - e^{-ity}) d\mu(t) - \int (e^{i(t-x)y} - e^{i(x-t)y}) \, d\mu(t)\right\} dy,$$

that is,

$$\int_{-r}^{r} = 2\int_{0}^{r} \left\{ y^{-1} \int \sin(ty) \, d\mu(t) - y^{-1} \int \sin(tx - ty) \, d\mu(t)\right\} dy.$$

Since $(\sin z)/z$ remains continuous at $z = 0$, and since $r < \infty$ and $\int |d\mu(t)| < \infty$, the order of integrations can be interchanged here. Thus

$$\int_{-r}^{r} = 2\int \left\{ \int_{0}^{r} y^{-1} \sin(ty) \, dy - \int_{0}^{r} y^{-1} \sin(tx - ty) \, dy\right\} d\mu(t)$$

or, since $\mathrm{si}(u) = \int_{0}^{u} v^{-1} \sin v \, dv$,

$$\int_{-r}^{r} = 2\int \mathrm{si}(rt) \, d\mu(t) - \int \mathrm{si}(rt - rx) \, d\mu(t).$$

Since this is identical with the integral on the right of (185), the assertion of (185) follows from (181) and from the case $x = 0$ of (181).

Similarly, if x in (10) is replaced by t, the integral on the right of (186) appears in the form

$$\int_{-r}^{r} = \int_{0}^{r} \left\{ e^{-ixy} \int e^{ity} \, d\mu(t) + e^{ixy} \int e^{-ity} \, d\mu(t)\right\} dy,$$

which can be rearranged into

$$\int_{-r}^{r} = \int \left\{ \int_{0}^{r} (e^{i(t-x)y} + e^{i(x-t)y}) \, dy\right\} d\mu(t) = \int \left\{ \int_{0}^{r} 2 \cos (t - x)y \, dy \right\} d\mu(t).$$

Since the interior integral, $\{\quad\}$, is 2r times the integral on the left of (183), it follows that (186) is identical with (183).

52. For a given distribution μ, let ${}_{-1}\mu$, as before (17), denote the distribution given by ${}_{-1}\mu(x) = 1 - \mu(-x)$ at all of its continuity points. Then an x is in the point spectrum of ${}_{-1}\mu$ if and only if $-x$ is in the point spectrum of μ. It follows therefore from (72), (68) and from the definition of a vectorial sum, that the point spectrum of μ is vacuous if and only if $x = 0$ is not in the point spectrum of ${}_{-1}\mu * \mu$. Hence, if μ and x in (186) are replaced by ${}_{-1}\mu * \mu$ and $x = 0$ respectively, it is seen from (54) that the condition

$$\int_0^r |\mu^o(y)|^2 \, dy = o(r) \text{ as } r \to \infty \tag{187}$$

is necessary and sufficient in order that the distribution μ be continuous in the sense of §4. It follows that

$$\int_0^r |\mu^o(y)| \, dy = o(r) \text{ if and only if } \mu \text{ is continuous.} \tag{188}$$

In fact, it is clear from (11) that the o-condition of (188) implies the o-condition of (187); while the converse inference is clear from Schwarz's inequality.

A corollary of (188) is that

(189) if $\lim\limits_{y\to\pm\infty} \mu^{o}(y) = 0$, then μ is continuous.

But it will be shown in § 63 that

(190) the converse of (189) is false.

On the other hand, a weakened form (189) has a converse, represented by the Riemann-Lebesgue lemma, according to which

(191) if μ is absolutely continuous, then $\lim\limits_{y\to\pm\infty} \mu^{o}(y) = 0$.

But it will be shown in § 64 that

(192) the converse of (191) is false.

As to (191), it is clear from (10) that, if μ is absolutely continuous and has a density $\mu'(x)$ which is constant on an interval and vanishes outside this interval, then $\mu^{o}(y) = O(|y|^{-1}) = o(1)$ as $y \to \pm\infty$. Hence, the same is true if $\mu'(x)$ is a step-function, having a finite number of intervals of constancy and vanishing outside a bounded interval, say $-r \leqq x \leqq r$. If μ is any absolutely continuous function, and if $\mu^{o}(y)_r$ is defined so as in § 10, then

$$\mu^{o}(y)_r = \int_{-r}^{r} e^{ixy}\mu'(x)\,dx; \text{ hence} \left|\mu^{o}(y)_r - \int_{-r}^{r} e^{ixy} f(x)\,dx\right| \leqq \int_{-r}^{r} |\mu'(x) - f(x)|\,dx,$$

where f is any L-integrable function on the interval $-r \leqq x \leqq r$. Since $\mu'(x)$ is L-integrable, there exists for every $\varepsilon > 0$ a step-function $f = f_\varepsilon$ having a finite number of intervals of constancy and such that the integral on the right of the inequality of the last formula line becomes less than ε. Finally, it is clear from § 10 that there exists for every $\varepsilon > 0$ an $r = r_\varepsilon$

such that $\left|\overset{o}{\mu}(y) - \overset{o}{\mu}(y)_r\right| < \varepsilon$ for every y. This, when combined with the inequality in the last formula line, gives,

$$\left|\overset{o}{\mu}(y)\right| < \varepsilon + \varepsilon + \left|\int_{-r}^{r} e^{ixy} f(x)\,dx\right| \text{for every } y.$$

Since the last integral was seen to be $O(|y|^{-1}) = o(1)$ as $y \to \pm\infty$, the proof of (191) is complete.

There is no explicit criterion necessary and sufficient for transforms $\overset{o}{\mu}(y)$ belonging to absolutely continuous distributions $\mu(x)$. On the other hand, it is seen from (10) and from Plancherel's completeness theorem, that a transform does or does not belong to an absolutely continuous distribution $\mu(x)$ having a density $\mu'(x)$ which is of class (L^2) on the line $-\infty < x < \infty$ according as $\overset{o}{\mu}(y)$ is or is not of class (L^2) on the line $-\infty < y < \infty$.

In particular, the existence of an $\varepsilon > 0$ satisfying $\overset{o}{\mu}(y) = O(|y|^{-\frac{1}{2}-\varepsilon})$ as $y \to \pm\infty$ is sufficient in order that μ be absolutely continuous. It can be shown that this is false for every $\varepsilon < 0$.

53. Thus it is of interest that

$$(193) \qquad \int\left|\overset{o}{\mu}(y)\right| dy < \infty$$

turns out to be sufficient for the absolute continuity of μ. In fact, (193) is sufficient in order that $\mu(x)$ should possess a density $\mu'(x)$ which is continuous for $-\infty < x < \infty$ and tends to 0 as $x \to \pm\infty$.

In order to see this, it is sufficient to observe that, if (193) is satisfied, then integral (185) and the integral $\int e^{-ixy} \mu^{o}(y)\, dy$, which results by differentiation of (185) beneath the integral sign, are uniformly convergent for $-\infty < x < \infty$ and represent therefore continuous functions of x. Hence it is clear from (185) that, if (193) is satisfied, (10) can be written in the form

$$(194) \qquad \mu^{o}(y) = \int e^{ixy} \mu'(x)\, dx,$$

where

$$(195) \qquad \mu'(x) = \frac{1}{2\pi} \int e^{-iyx} \mu^{o}(y)\, dy.$$

Finally, a repetition of the proof of (191) shows that (193) and (195) imply that $\mu'(x) \to 0$ as $x \to \pm\infty$.

It may be mentioned that, since $\mu(x)$ is supposed to be a monotone function satisfying (8), the conditions under which the "Fourier inversion", (195), of (194) was just established are symmetric in x and y; in fact, (193) is paralleled by $\int |\mu'(x)|\, dx < \infty$, since $\int |\mu'(x)|\, dx = \int d\mu(x) = 1$. Furthermore, it was shown in § 10 that $\mu^{o}(y)$ is always continuous. Finally, $\mu^{o}(y) \to 0$ as $y \to \pm\infty$, by (191).

If the assumption (193) is replaced by

$$(196) \qquad \int |y|^{k} |\mu^{o}(y)|\, dy < \infty,$$

where k is a non-negative integer, then the integrals which result by m-fold formal differentiation of (195) are uniformly convergent for $-\infty < x < \infty$, if $m \leq k$. Consequently, (196) is sufficient in order that $\mu(x)$ should have a continuous density possessing continuous derivatives of any order not exceeding k. Since all these derivatives can be obtained by differentiating (195) beneath the integral sign, it follows by a repetition of the proof of

(191), that all these derivatives tend to 0 as $x \rightarrow \pm\infty$.

The existence of some $q > 0$ satisfying

$$\mu^{o}(y) = O(|y|^{-k-1-q}) \text{ as } y \rightarrow \pm\infty \tag{197}$$

is sufficient for (196). In particular, the existence of some $p > 0$ satisfying

$$\mu^{o}(y) = O(\exp(-|y|^{p})) \text{ as } y \rightarrow \pm\infty \tag{198}$$

is sufficient in order that (196) be satisfied for every k; so that $\mu'(x)$ then has derivatives of arbitrarily high order for $-\infty < y < \infty$. But the formal Taylor series of $\mu'(x)$ at $x = 0$, i.e., the power series

$$\sum_{n=0}^{\infty} \mu^{(n+1)}(0)\, x^{n}/n! = (2\pi)^{-1} \sum_{n=0}^{\infty} (-1)^{n} \int y^{n}\, \mu^{o}(y)\, dy\, x^{n}/n!, \tag{199}$$

can then be divergent for every $x \neq 0$. In fact, it will be seen in § 54 that there exists for every $p < 1$ a distribution μ satisfying (198) but rendering the series (199) divergent for every $x \neq 0$.

However, if (198) is satisfied for $p = 1$, and, more generally, if there exists a positive q satisfying

$$\mu^{o}(y) = O(\exp(-q|y|)) \text{ as } y \rightarrow \pm\infty, \tag{200}$$

then $\mu'(x)$ is regular-analytic for $-\infty < x < \infty$. In fact, there must then exist in the strip $|v| < q$ of the w-plane, where $w = x + iv$, a regular-analytic function which is identical with $\mu'(x)$ on the real axis. In addition, this regular-analytic continuation must then be bounded on every strip $|v| < r$, if $r < q$. All of this follows by observing that the integral which results by replacing x by $w = x + iv$ in (195) is uniformly convergent, and represents a bounded function, on the strip $|v| < r$, if r is less than the constant q for which (200) is supposed.

If (198) is satisfied by some $p > 1$, then (200) is satisfied by every q, and so $\mu'(x)$ must then possess an analytic continuation which is an entire function. In addition, the latter must then be bounded on every fixed strip parallel to the real axis. Finally, the order of this entire function cannot exceed $p/(p - 1)$. This is readily verified from Stirling's formula, since (198), where by assumption, $p > 1$ implies that the coefficient of x^n in (199) is majorized by a constant multiple of

$$\int_0^\infty y^n \exp(-y^p)\, dy = p^{-1}\Gamma([n + 1]/p)/n!. \tag{201}$$

54. As an illustration, consider the distribution $\mu = \mu_p$ for which the transform is exactly

$$\mu_p^{\,o}(y) = \exp(-|y|^p), \tag{202}$$

provided that μ_p exists. This proviso is necessary, since the function of x which results if (202) is substituted into (195) can attain negative values and, therefore, fail to be a density.

Actually, the distribution μ_p cannot exist if $p > 2$. In fact, (202) implies that $\mu_p^{\,o}(y) - 1 \sim -|y|^p$ as $y \rightarrow 0$. Hence, the o-condition of (20) is satisfied whenver $p > 2$. Consequently, (20) states that, if μ_p exists for a $p > 2$, then $\mu_p = \varepsilon$. But then $\mu_p^{\,o}(y) = 1$ for all y, by (15). And this contradicts (202), the limiting case $p = 0$ being excluded by $p > 2$.

It will from now on be assumed that $0 < p \leqq 2$. The existence of the distribution μ_p for every such p will be proved in § 56. Let this be granted for the present.

Since (202) is an even function, the coefficient of $(2\lambda)^{-1} x^n$ in (199) is 0 or $2i^n$ times the integral (201) according as n is odd or even. Accordingly, the Taylor series at $x = 0$ for the density of μ_p is

$$(203) \qquad \sum_{n=0}^{\infty} \mu_p^{(n+1)}(0)x^n/n! = \frac{1}{\pi p}\sum_{n=0}^{\infty}(-1)^n \frac{\Gamma([2n+1]/p)}{\Gamma(2n+1)}x^{2n}$$

According to Stirling's formula, the radius of convergence of (203) is 0 or ∞ according as $p < 1$ or $p > 1$. This agrees with the end of § 53.

In the limiting case $p = 1$, the radius of convergence of (203) is neither 0 nor ∞ and the density is seen to become

$$(204) \qquad \mu_1'(x) = \frac{1}{\pi}\,\frac{1}{1 + x^2}$$

(for all x). This rational function becomes singular at the points $\pm i$, points having the distance 1 from the origin. Since (200) is now satisfied by $q = 1$, it follows that the conclusions drawn from (200) in § 53 cannot in general be improved.

If $p = 2$, then, since $\Gamma(n + \frac{1}{2}) = \pi^{\frac{1}{2}}2^{-2n}\,\Gamma(2n + 1)/n!$,

$$(205) \qquad \mu_2'(x) = \tfrac{1}{2}\pi^{-\frac{1}{2}}\exp(-|\tfrac{1}{2}x|^2).$$

If the case $p = 2$ of (202) is compared with (154), it is seen from (10) and (17) that (205) is equivalent to

(205)$_{\text{bis}}$ $$\theta'(x) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2), \text{ where } \theta^0(y) = \exp(-\tfrac{1}{2}y^2).$$

Since substitution of (202) into (195) shows that

(206) $$\mu'_p(x) = \pi^{-1}\int_0^\infty e^{-t^p} \cos(xt)\,dt,$$

the density $\mu_p'(x)$ is an even function for every p. It can therefore be assumed that $x \geqq 0$.

A partial integration of (206) gives

$$\mu_p'(x) = \pi^{-1}p \int_0^\infty e^{-t^p} t^p S(xt)\,dt, \text{ where } S(z) = (\sin z)/z.$$

Hence, if t is replaced by $(t/x)^{1/p}$ for a fixed $x > 0$, it follows that

(207) $$\mu_p'(x) = \pi^{-1}\alpha_p(1/x^p)/x^{p+1}, \qquad (x^p > 0),$$

where $\alpha_p(z)$ is an abbreviation for

(208) $$\alpha_p(z) = \int_0^\infty e^{-zt} \sin(t^{1/p})\,dt.$$

If $p < 1$, the Laplace transform (208) is convergent at $z = 0$ and has there the value $\alpha_p(0) = \Gamma(p)\sin(\frac{1}{2}\pi p)$; so that, since the integral analogue of Abel's continuity theorem ensures that $\alpha_p(+0) = \alpha_p(0)$, there results the limit relation $\alpha_p(\varepsilon) \to \Gamma(p)\sin(\frac{1}{2}\pi p)$ as $\varepsilon \to +0$. Actually, it is easily verified from (208), for instance by residue calculus, that (in accordance with Heaviside's ideas) this limit relation remains valid for the case $p > 1$, i.e., for the case in which the integral $\alpha_p(0)$ is divergent. Consequently, from (207),

(209) $$x^{p+1}\mu_p'(x) \to \pi^{-1}\Gamma(p)\sin(\tfrac{1}{2}\pi p) \text{ as } x \to \infty.$$

If $p < 2$, the constant on the right of (209) is positive, and so (209) implies that $\int |x|^{p-\varepsilon} d\mu_p(x) < \infty$ if $0 < p - \varepsilon < p$ but $\int |x|^p d\mu_p(x) = \infty$ (if $p = 2$, then (204) shows that the situation is quite different). It follows, in particular, that the second moment, $\int x^2 d\mu_p(x)$, is ∞ for every $p < 2$. This is implied by the easier results of § 12 also, since, if $p < 2$, it is obvious from (202) that there cannot exist two constants, c_0 (=1) and c_1, satisfying (34) for $n = 1$.

55. In order to discuss the analytic character of μ_p', it can be assumed that $p < 1$, since μ_p' is the rational function (204) or the transcendental entire function (203) according as $p = 1$ or $p > 1$.

If $p < 1$, then, while $\mu_p'(x)$ has derivatives of arbitrarily high order for $-\infty < x < \infty$, the series (203) diverges for every $x \neq 0$. More than this will now become understandable from (207). In fact, it will be shown that $\alpha_p(z)$ is a transcendental entire function, if $p < 1$. Needless to say, the Laplace expansion (208) of this entire function is divergent on the left of the imaginary axis. But it turns out that, if $p < 1$, the function defined by (208) on the right of the imaginary axis possesses the analytic continuation

$$(210) \quad \sum_{n=0}^{\infty} \alpha_p^{(n)}(0) z^n/n! = p \sum_{n=0}^{\infty} (-1)^n \Gamma([n+1]p) \sin(\tfrac{1}{2}\pi [n+1]p)\, z^n/n!.$$

If $p = \frac{1}{2}$, then (210) can be contracted, corresponding to (205), into

$$(211)\qquad 2\alpha_{\frac{1}{2}}(z) = (\tfrac{1}{2}\pi)^{\frac{1}{2}} e^{-(\frac{1}{2}z)^2} - \sum_{n=0}^{\infty} (-1)^n n! z^{2n+1}/(2n+1)!\ .$$

In order to prove (210), where $p < 1$, let q and r be positive constants and let

$$\int_C f(z)\,dz, \text{ where } f(z) = z^{q-1} e^{iz - rz^p},\ (z^b > 0 \text{ for } z > 0)$$

be applied to the contour C consisting of the two segments and two circular arcs

$$z = t,\ z = it;\quad z = \varepsilon e^{i\phi},\ z = Re^{i\phi},$$

where $0 < \varepsilon < t < R < \infty$ and $0 \leqq \phi \leqq \frac{1}{2}\pi$. Since $q > 0$, $r > 0$ and $0 < p < 1$, it is readily ascertained that the contribution of the two circular arcs to the contour integral tends to 0 as $\varepsilon \longrightarrow 0$ and $R \longrightarrow \infty$. Hence, the integrals of $f(t)$ and of $if(it)$ over the half-line $0 < t < \infty$ have a common value. Thus, by definition of $f(z)$,

$$(212)\quad \int_0^\infty e^{-rt^p} t^{q-1} e^{it}\,dt = e^{\frac{1}{2}\pi q i} \int_0^\infty e^{-t-rt^q\cos\frac{1}{2}\pi p}\, t^{q-1} e^{-irt^q\sin\frac{1}{2}\pi p}\,dt.$$

Since $p > 0$, the n-th derivative of the regular-analytic function represented by (208) in the half-plane to the right of the imaginary axis can be obtained by differentiating (208) beneath the integral sign. Thus, if $z = r > 0$,

$$\alpha_p^{(n)}(r) = (-1)^n \int_0^\infty e^{-rt} t^n \sin(t^{1/p})\,dt \equiv (-1)^n p \int_0^\infty e^{-rt^p} t^{(n+1)p-1} \sin t\,dt,$$

if t is replaced by t^p. Hence, by the imaginary part of (212),

$$a_p^{(n)}(r) = (-1)^n p \int_0^\infty e^{-t-rt^p\cos\frac{1}{2}\pi p}\, t^{(n+1)p-1} \sin(\tfrac{1}{2}\pi[n+1]p - rt^p \sin\tfrac{1}{2}\pi p)\, dt,$$

where $(n + 1)p = q$ and $0 < r <$. Since $0 < p < 1$, it follows that

$$(213)\quad |a_p^{(n)}(r)| < p\int_0^\infty e^{-t} t^{(n+1)p-1} dt \equiv p\,\Gamma([n+1]p) \text{ for } 0 < r < \infty,$$

and that $a_p^{(n)}(r)$ therefore tends, as $r \longrightarrow +0$, to the limit

$$(-1)^n p \int_0^\infty e^{-t} t^{(n+1)p-1} \sin(\tfrac{1}{2}\pi[n + 1]p)\, dt.$$

In view of Rolle's theorem, this implies that the derivative $a_p^{(n)}(0)$, when thought of as a right-hand derivative, exists and has the value represented by the last formula line, i.e., by the value formally assigned by the assertion (210).

Finally, since (208) represents a regular-analytic function in the half-plane to the right of the imaginary axis, $a_p(z)$ can be developed in the vicinity of every $z = r > 0$ into a power series,

$$a_p(z) = \sum_{n=0}^{\infty} a_p^{(n)}(r)(z - r)^n/n!,$$

having a radius of convergence not less than r. But (213) shows that this power series is dominated by one which, in view of Stirling's formula, is convergent in the whole $(z - r)$-plane, since $0 < p < 1$. Consequently, $a_p(z)$ is an entire function. Hence, (210) represents $a_p(z)$ for every z.

If $p > 1$, then (210) diverges for every $x \neq 0$, as (203) does in the case $p < 1$. If $p = 1$, then (210) converges in the circle $|z| < 1$ to the function $1/(1 + z^2)$; so that (207) becomes the standard functional equation of the rational function (204). Finally, (203) and (210) are transcendental entire functions of the respective orders $(1 - p^{-1})^{-1}$, $(1 - p)^{-1}$ in the complementary cases $p > 1$, $p < 1$.

If $p > 1$, then (210) diverges for every $x \neq 0$, as (203) does in the case $p < 1$. If $p = 1$, then (210) converges in the circle $|z| < 1$ to the function $1/(1 + z^2)$; so that (207) becomes the standard functional equation of the rational function (204). Finally, (203) and (210) are transcendental entire functions of the respective orders $(1 - p^{-1})^{-1}$, $(1 - p)^{-1}$ in the complementary cases $p > 1$, $p < 1$.

56. What remains to be shown is the existence of a distribution μ_p satisfying (202), where $0 < p \leqq 2$.

It turns out that, in order to prove the existence of μ_p for a given p, it is sufficient to ascertain the existence of some distribution $\lambda = \lambda_p$ satisfying the relation

$$(214) \qquad \overset{o}{\lambda_p}(y) = 1 - y^p + o(|y|^p) \text{ as } y \longrightarrow 0$$

(this relation is satisfied by (202), if $\lambda_p = \mu_p$ is known to exist, but it can easier be ascertained for some λ_p than in the explicit case required). In fact, since (214) is equivalent to $\log \lambda_p^o(y) = - y^p + o(|y|^p)$, it implies that, as $n \longrightarrow \infty$, the function $n \log \lambda_p^o(n^{-1/p}y)$ of y tends to $-|y|^p$ uniformly on every fixed bounded y-interval. Hence, if the existence of a distribution λ_p satisfying (214) is known, it follows, in exactly the same say as at the end of § 44, that the n-th power of $\lambda_p^o(n^{-1/p}y)$ is the transform of a distribution, and that the latter tends, as $n \longrightarrow \infty$, to a distribution μ_p satisfying (202).

In accordance with the general remark made at the beginning of § 13, the satisfaction or violation of (214) by a given distribution $\lambda = \lambda(x)$ depends, in the main, on the behavior of $\lambda(x)$ as $x \rightarrow \pm\infty$. It turns out that the behavior necessary for (214) is typified by the distribution $\lambda(x)$ having the density

$$(215) \qquad \lambda_p'(x) = C_p \,/(1 + x^2)^{\frac{1}{2}p+\frac{1}{2}}, \qquad (C_p = \text{const.}),$$

if either $0 < p < 1$ or $1 < p < 2$. The remaining cases, $p = 1$ and $p = 2$, are taken care of by (204) and (205). It is understood that the numerical value of the constant C_p is determined by the condition $\int d\lambda_p(x) = 1$. According to (215) and (10), the transform is

$$(216) \qquad \lambda_p^{\,o}(y) = 2C_p \int_0^\infty (1 + x^2)^{-\frac{1}{2}p-\frac{1}{2}} \cos (xy)\, dx,$$

and can therefore be expressed in terms of the cylinder function

$$(216 \text{ bis}) \qquad (\tfrac{1}{2}z)^q K_q(z) = \pi^{-\frac{1}{2}}\, \Gamma(q + \tfrac{1}{2}) \int_0^\infty (1 + t^2)^{-q-\frac{1}{2}} \cos (zt)\, dt; \quad q > 0.$$

If x is replaced by xy for a fixed $y \neq 0$, and if it is assumed, without loss of generality, that $|y| = y$, it is seen that the ratio of $1 - \lambda_p^{\,o}(y)$ to y^p is a positive constant multiple of

$$y^{-1}y^{-p} \int_0^\infty (1 + x^2/y^2)^{-\frac{1}{2}p-\frac{1}{2}} \sin^2 \tfrac{1}{2}x \, dx.$$

Hence, in order to show that (214) becomes satisfied upon a suitable choice of the unit of length on the y-axis (or, what according to (17), is the same thing, upon the reciprocal choice of the unit of length on the x-axis), it is sufficient to verify that the expression in the last formula line tends to a positive limit as $y \longrightarrow +0$. But this becomes clear in both cases, $0 < p < 1$ and $1 < p < 2$, if the integration range $0 \leqq x < \infty$ is split at $x = 1$.

This completes the proof for the existence of the distributions μ_p implicitly defined by (202), where $0 < p \leqq 2$.

56 bis. It is by no means obvious that all those distributions μ_p exist for which the condition $p \leqq 2$, imposed by the trivial restriction (20), is not violated (cf. the beginning of § 54). In order to see this, suppose that (202) is replaced by the assignment

$$\mu_p^{\,0}(y) = \mathrm{Max}\,(0,\ 1 - |y|^p), \tag{202 bis}$$

where $p > 0$. Then, since the expression on the right of (202 bis) is $1 - |y|^p$ or 0 according as $|y| \leqq 1$ or $|y| \geqq 1$, the restriction (20) imposes again the condition $p \leqq 2$ only. But it turns out that even the range $1 < p \leqq 2$ must now be excluded, since otherwise the distribution μ_p assigned by (202 bis) does not exist.

Suppose that this distribution exists for a given p. Then, since (202 bis) vanishes for large $|y|$, it follows from the criterion found at the beginning of § 53, that μ_p has a density represented by (195). But substitution of (202 bis) into (195) gives

$$\mu_p{}'(x) = \pi^{-1} \int_0^1 (1 - t^p) \cos(xt)\, dt. \tag{206 bis}$$

It follows therefore by partial integration that

$$\mu_p{}'(x) = \pi^{-1} p\, \alpha_p(x)/x \text{ if } x > 0, \tag{207 bis}$$

where $\alpha_p(x)$ is an abbreviation for

$$\text{(208 bis)} \qquad \alpha_p(x) = \int_0^1 t^{p-1} \sin(xt)\, dt.$$

Since (206 bis) is an even continuous function, (207 bis) shows that (206 bis) is non-negative for $-\infty < x < \infty$ if and only if (208 bis) is non-negative for $0 < x < \infty$. But (206 bis) is the density of a distribution. Accordingly, the distribution μ_p assigned by (202 bis) exists if and only if p is such as to make the function (208 bis) non-negative for every $x > 0$. And it turns out that this is the case if and only if $p < 1$.

In fact, (208 bis) can be written in the form

$$\text{(209 bis)} \qquad x^p \alpha_p(x) = \int_0^x t^{p-1} \sin t\, dt, \qquad \text{where } x > 0.$$

If $n = 1, 2, \ldots$, the absolute value of the contribution of the interval $(n - 1)\pi \leqq t \leqq n\pi$ to the integral (209) is an increasing or a decreasing function of n according as the function t^{p-1} of t is increasing or decreasing on the half-line $0 < t < \infty$, i.e., according as $p > 1$ or $0 < p < 1$. Moreover, since $(-1)^{n-1} \sin t$ is positive in the interior of the n-th interval, the n-th contribution has the same sign as $(-1)^{n-1}$, and is therefore positive for the first value of n. Consequently, the function (209 bis) becomes negative for certain values of $x > 0$ or is positive for every $x > 0$ according as $p > 1$ or $0 < p < 1$. Since (209 bis) becomes the non-negative function $1 - \cos x$ in the limiting case $p = 1$, the proof is complete.

It is also seen that the case $p = 1$ represents the elementary distribution denoted by μ_1 at the end of § 7. If $0 < p < 1$, then (209 bis) shows that $x^p \alpha_p(x) \longrightarrow \Gamma(p) \sin(\frac{1}{2}\pi p)$ as $x \longrightarrow \infty$, which means, by (207 bis), that (209) becomes correct if $\Gamma(p)$ on the right of (209) is replaced by $p\,\Gamma(p) = \Gamma(p + 1)$.

Examples

57. For every $r > 0$, let $\beta_r = \beta_r(x)$ denote the distributions represented by the step-function which is 0, $\frac{1}{2}$ or 1 according as $-\infty < x < -r$, $-r < x < r$ or $r < x < \infty$. Thus $\beta_r{}^o(h) = \cos(ry)$, by (10). Hence, if $r_1, r_2, \ldots$ is a sequence of positive numbers, the criterion of (95) shows that the infinite convolution $*\beta_{r_n}$ is convergent if and only if the product $\prod \cos(r_n y)$ is uniformly convergent on every fixed bounded y-interval; in which case the transform of the distribution

$$\mu = *\beta_{r_n} \tag{217}$$

is

$$\mu^o(y) = \prod \cos(r_n y). \tag{218}$$

Since $\cos t = 1 - \frac{1}{2}t^2 + o(t^2)$ as $t \to 0$, it is clear that $\prod \cos(r_n y)$ is uniformly convergent on every fixed y-interval if and only if the positive constants are so chosen that

$$\sum r_n{}^2 < \infty. \tag{219}$$

It will always be assumed that (219) is satisfied, i.e., that (217) defines a distribution μ. Then the product (218) is absolutely convergent for every u, i.e., the convergent convolution (217) must converge absolutely (this agrees with the remark made at the end of §32). It follows therefore from (219) that the notation in (217) can be so chosen that

$$r_n \geqq r_{n+1} \text{ for every } n \qquad (r_n > 0). \tag{220}$$

Since the spectrum of β_r consists of the two points $x = \pm r$, it is clear from the description (98) of the spectrum of an arbitrary convergent convolution, that a point $x = s$ is in the spectrum $S(\mu)$ of the distribution defined by (217) if and only if there exists to s and to every n a decision of the alternative $\pm r_n$ such that s becomes the sum of a convergent series of the form

(221) $$s = \sum \pm r_n .$$

This implies that $S(\mu)$ is a bounded set on the x-axis if and only if (219) is refined to

(222) $$\sum r_n < \infty$$

and that $s = \sum r_n$ then is the greatest, and $s = -\sum r_n$ the least, value contained in $S(\mu)$. If, on the other hand, $\sum r_n = \infty$, then, since $r_n \to 0$ in virtue of (219), it is clear that there exists to every real number x and to every n a decision on the alternative $\pm r_n$ in such a way that x appears in the form $\sum \pm r_n$ of a (conditionally) convergent series. Accordingly, $S(\mu)$ is a bounded set or the whole x-line according as the series (221) is or is not absolutely convergent.

58. It turns out that

(223) μ is a continuous distribution for every choice of the series (219) which defines the convergent convolution (217).

According to the criterion (188) or to its equivalent formulation (187), the assertion of (223) is that (218) satisfies the estimate

(224) $$\int_0^t \prod |\cos(r_n y)| dy = o(t) \text{ or } \int_0^t \prod \cos^2(r_n y) dy = o(t) \text{ as } t \to \infty .$$

Since every factor of the infinite products $\prod$ is at most 1 for every y, it is sufficient to verify (224) for the case in which the sequence $r_1, r_2, \ldots$ is replaced by any of its subsequences, say by its subsequence $r_1', r_2', \ldots$. Actually, if $r_1'', r_2'', \ldots$ denote the remaining terms of the sequence $r_1, r_2, \ldots$, and if μ' and μ'' are the convolutions which correspond to these two complementary subsequences, then $\mu = \mu' * \mu''$, by (217), (219) and (92). It follows therefore from (75) that it is sufficient to prove (223) or (224) for the case where (217) is replaced by the convergent convolution μ' belonging to some subsequence $r_1', r_2', \ldots$ of $r_1, r_2, \ldots$. However, it will be seen in § 60 that the continuity of the distribution μ' is obvious if the selection of the subsequence $r_1', r_2', \ldots$ defining μ' is sufficiently retarded, namely such that r_n' becomes greater than the value of the remainder series $r_{n+1}' + r_{n+2}' + \ldots$, where $n = 1, 2, \ldots$. If this is granted for the present, the truth of (223) follows for every distribution (217). Incidentally, this proof of (223) can be varied so as to be based on a corresponding selection in (224); cf. the end of § 64 below.

According to the definition at the beginning of § 57, the distributions β_r are purely discontinuous. It follows therefore from (148) that the distribution (217) is pure. This, when combined with (223), means that

(225) μ is either absolutely continuous or purely singular.

Another consequence of (223) is that

(226) $S(\mu)$ is a bounded perfect set or the whole x-axis according as $\sum r_n < \infty$ or $\sum r_n = \infty$.

In fact, $S(\mu)$ can, by § 4, be defined as the set of those points x in some vicinity of which the function $\mu(x)$ is not constant. This clearly implies that $S(\mu)$ must be perfect whenever μ is continuous. Hence, (226) follows from (223) and from what was shown at the end of § 57.

59. At the beginning of §36, there was considered a sequence of "spaces" $T_1, T_2, \ldots$ consisting of respective "points" $t_1, t_2, \ldots$. It was assumed that every T_n carries a measure ϕ_n satisfying (117). Finally, it was assumed, after (121), that every T_n carries a ϕ_n-measurable function $x_n = x_n(t_n)$ of the position.

An instructive realization of these requirements can be based on the distributions β_r. In fact, let every T_n consist of only two "points" t_n. Let these points be denoted by $t_n = +$ and $t_n = -$ for every n and let the ϕ_n-measure of either of these points be assigned to be $\frac{1}{2}$ for every n (so that (117) is satisfied). Finally, let the function $x_n(t_n)$ of the position $t_n = \pm$ on T_n be defined by placing $x_n(\pm) = \pm r_n$, where r_n is a positive number.

It is clear from the definition (114) of the ϕ_n-distribution of a function $x_n(t_n)$ on T_n that, in the present case, the ϕ_n-distribution of $x_n(t)$ on T_n is 0, $\frac{1}{2}$ or 1 according as $-\infty < x < -r_n$, $-r_n < x < r_n$ or $r_n < x < \infty$; that is, the distribution β_{r_n} defined at the beginning of § 57. It follows therefore from (138) that the series $\sum x_n(t_n)$, i.e., $\sum \pm r_n$, converges on the infinite product space $T_1 . T_2 . \ldots$ of the points

(227) $t = (t_1, t_2, \dots) = (\pm, \pm, \dots)$

in $\prod \phi_n$-measure if and only if the infinite convolution (217) is convergent, i.e., if and only if (219) is satisfied; in which case the distribution (217) is precisely the $\prod \phi_n$-distribution of the function

(228) $x(t) : \sum \pm r_n$

to which $\sum \pm r_n$ converges in $\prod \phi_n$-measure on the space $T = T_1 . T_2 . \; \dots$ of the points (227).

All of this sounds more abstract than it actually is. In fact, the infinite product space $T = T_1 . \; T_2 . \; \dots$, consisting of the points (227) and carrying the $\prod \phi_n$-measure, can be mapped on the interval $0 < t < 1$, consisting of real numbers t and carrying the Euclidean Lebesgue measure, as follows:

Let the two points, $t_n = \pm$, of T_n be thought of as representing the alternative which allots either 1 or 0 to the n-th place in the infinite dyadic expansion of a real number t satisfying $0 < t < 1$; so that

(229) $t = \sum t_n / 2^n$ and $t_n = \frac{1}{2}(1 \pm 1)$

replace the points $t = (\pm, \pm, \dots)$ and $t_n = \pm$ of $T = T_1 . T_2 . \; \dots$ and of T_n respectively. Except for those points of the interval $0 < t < 1$ which possess a finite dyadic expansion (and form, therefore, an enumerable set, and hence a set of Euclidean measure 0), this mapping of the infinite product space $T_1 . T_2 . \; \dots$ on the interval $0 \leqq t \leqq 1$ is one-to-one. Moreover, the mapping is measure-preserving, that is, the $\prod \phi_n$-measure of a subset of $T_1 . T_2 . \; \dots$ and the Euclidean measure of the image of this subset on the interval $0 \leqq t \leqq 1$ exist at the same time, and the two measures always have the same value. This follows from (229) and from the definition of the product measure $\prod \phi_n$ on $T = T_1 . T_2 . \ldots$, since the measure on T_n was assigned by $\phi_n(\pm) = \frac{1}{2}$.

Accordingly, if the terms ofthe series (228) are thought of as functions of the position (229) on the interval $0 \leqq t \leqq 1$, then the series (228) converges in Euclidean measure to a function $x(t)$ on the interval $0 \leqq t \leqq 1$ if and only if the condition, (219), for the convergence of (217) is satisfied; in which case the distribution (217) is the ϕ-distribution of $x(t)$ on $0 \leqq t \leqq 1$, where ϕ denotes the Euclidean measure.

Clearly, (228) and (229) can be written as

$$(230) \quad x(t): \quad \tfrac{1}{2} \sum (1 \pm 1) r_n/2^n, \text{ where } t = \tfrac{1}{2} \sum (1 \pm 1)/2^n; \quad (0 < t < 1).$$

If $[u]$ denotes the greatest integer not exceeding u, it is easily verified that (230) is equivalent to

$$(231) \qquad x(t) : \quad - \sum (-1)^{[2^n t]} r_n; \qquad 0 \leqq t \leqq 1.$$

In fact, the $[2^n t]$-th power of -1 is identical with $-(-1)^k$ for every t contained in the interior of the k-th subinterval obtained by dividing the interval $0 \leqq t \leqq 1$ into 2^n congruent subintervals ($k = 1, \ldots, 2^n$).

60. Suppose that not only (219) but also (222) is satisfied. Then the series (231) is uniformly convergent and represents, therefore, the function to which it converges in Euclidean measure on $0 \leqq t \leqq 1$. Thus

(231) can be replaced by

$$x(t) = -\sum (-1)^{[2^n t]} r_n, \text{ where } \sum r_n < \infty ; \qquad 0 \leq t \leq 1 . \tag{232}$$

In view of the identity mentioned after (231) the n-th term of the series (232) is discontinuous only at the points $t = k/2^n$ of the interval $0 \leq t \leq 1$. Hence, every term of the series (232), and therefore, by uniform convergence, the function $x(t)$ itself, is continuous at all those points t of the interval $0 < t < 1$ which do not possess a finite dyadic expansion, i.e., which cannot be represented in the form $t = t_{jm}$, where

$$t_{jm} = j/2^m, \text{ where } j = 1,3,\ldots,2^m - 1 \text{ and } m = 1,2,\ldots \tag{233}$$

The uniform convergence of the series (232) also implies that, at the points (233), the limits $x(t + 0)$ and $x(t - 0)$ exist and have the values which result if t is respectively replaced by $t + 0$ and $t - 0$ in each of the terms of the series (232). But it is clear from the identity mentioned after (231) that the difference between the right-hand and the left-hand limits of $(-1)^{[2^n t]}$ at a point $t = t_{jm}$ represented by (233) is 0, -2 or 2 according as $n < m$, $n = m$ or $n > m$. Consequently, from (232),

$$x(t_{jm} + 0) - x(t_{jm} - 0) = -(\sum_{n=1}^{m-1} 0 - 2r_m + \sum_{n=m+1}^{\infty} 2r_n).$$

This can be written in the form

$$x(t_{jm} + 0) - x(t_{jm} - 0) = 2(q_{m-1} - 2q_m), \tag{234}$$

if q_m is an abbreviation for

$$q_m = \sum_{n=m+1}^{\infty} r_n, \qquad q_{m-1} - q_m = r_m > 0; \qquad (q_0 = \sum r_n). \tag{235}$$

61. Of particular interest is the case of those series (222) for which all the jumps (234) are positive:

(236) $q_m < \frac{1}{2}q_{m-1}$, i.e. $q_m < r_m$, for every m,

cf. (235). Then the function $x(t)$ is monotone on the interval $0 \leqq t \leqq 1$. In fact, whether (236) is or is not satisfied, the n-th term of the series (232) is constant on each of the 2^n intervals mentioned at the end of §59. On the other hand, all the contributions of all the jumps of these step-functions to the function (232) are represented by (233) and (234). Hence, if the function (232) is increasing at all of its discontinuity points, i.e., if each of its jumps (234) is positive, then the function (232) is non-decreasing at every point of the interval $0 \leqq t \leqq 1$. Furthermore, the function then is nowhere constant, since the points (233), at which its jumps occur, are dense on the interval $0 \leqq t \leqq 1$. Accordingly,

(237) $x(t)$ is an increasing function on the interval $0 \leqq t \leqq 1$, if (236) is satisfied.

For a fixed m, there are 2^{m-1} points (233), and the jump of $x(t)$ is given by (234) at each of these points. Since, as observed after (232), the function $x(t)$ is continuous except at the points (233), it follows that the sum of all jumps of the monotone function $x(t)$ on the interval $0 \leqq t \leqq 1$ is $\sum 2^{m-1}2(q_{m-1} - 2q_m)$, where m runs, as in (233), through all positive integers. Obviously, this sum can be contracted into $2q_0 - a$, if a is an abbreviation for

(238) $$\lim_{n \to \infty} 2^{n+1} q_n.$$

This implies, in particular, that the assumption (236), which, via (237), underlies this calculation, necessitates the existence of the limit (238).

It is clear from (232) that $x(0) = -q_0$ and $x(1) = q_0$, where $q_0 = \sum r_n$. Hence, the total variation of the monotone function $x(t)$ on the interval $0 \leqq t \leqq 1$ is $2q_0$. Since the sum of all jumps of $x(t)$ was seen to be $2q_0$ minus the value of the limit (238), it follows by subtraction that (238) represents the total variation of the continuous component of the increasing function $x(t)$ on the interval $0 \leqq t \leqq 1$.

It will now be shown that

(240) $$\text{meas } S(\mu) = \lim_{n \to \infty} 2^{n+1} q_n \quad \text{in the case (236),}$$

where meas $S(\mu)$ denotes the Euclidean measure of the spectrum, $S(\mu)$, of the distribution (217). It will also be proved that

(241) $S(\mu)$ is nowhere dense in the case (236).

First, the result formulated before (230), when combined with the definition (114), can be expressed by saying that, even if only (219) is assumed, $\mu(x)$ is identical with the Euclidean measure of the set of those points of the interval $0 \leqq t \leqq 1$ at which the value of the function $x(t)$ is less than x, where x is any fixed real number. Hence, if the assumption, (236), of (237) is satisfied, then $\mu(x)$ is identical with the length of that portion of the interval $0 \leqq t \leqq 1$ on which the monotone function $x(t)$ is less than x. Since, as observed above, $x(0) = -q_0$ and $x(1) = q_0$, it follows, in particular, that $\mu(x) = 0$ for $-\infty < x < -q_0$ and that $\mu(x) = 1$ for $q_0 < x < \infty$.

Now consider the remaining interval, $-q_0 \leqq x \leqq q_0$, on which the interval $0 \leqq t \leqq 1$ is mapped by the monotone increasing function $x = x(t)$; it being understood that, if t is a discontinuity point of $x(t)$, then the subinterval $x(t - 0) < x < x(t + 0)$ of the interval $-q_0 < x < q_0$ does not contain image points of points of the interval $0 \leqq t \leqq 1$. Since $\mu(x)$ is the length of that portion of the interval $0 \leqq t \leqq 1$ on which the monotone function $x(t)$ is less than x, it follows that the function $\mu(x)$, where $-q_0 \leqq x \leqq q_0$, can be thought of as the inverse of the function $x(t)$, where $0 \leqq t \leqq 1$; it being understood that the jumps of the increasing function $x(t)$ are mapped on intervals of constancy of the non-decreasing function $\mu(x)$.

The points, (233), at which $x(t)$ has a jump, (234), are dense on the interval $0 \leqq t \leqq 1$. Consequently, the inverse function, $\mu(x)$, has no jump. Since the assumption (236) can, by (235), be written in the form $r_n > r_{n+1} + r_{n+2} + \ldots$, where $n = 1,2,\ldots$, this implies the fact anticipated in the proof of (223), if $r_1, r_2, \ldots$ is identified with the sequence $r_1', r_2', \ldots$ considered after (224).

According to §4, the spectrum $S(\mu)$ consists of those points x in some neighborhood of which the function μ is not constant. It follows therefore from the above interpretation of μ as the inverse of the function $x(t)$, that $S(\mu)$ consists of those points x of the interval $-q_0 \leqq x \leqq q_0$ which are not contained in any of the subintervals $x(t + 0) < x < x(t - 0)$, where t denotes any of the discontinuity points, (233), of $x(t)$. Consequently, the Euclidean measure of $S(\mu)$ results if the sum of all jumps (234) is subtracted from the length, $2q_0$, of the interval $-q_0 \leqq x \leqq q_0$.

This, when compared with the result found before (238), completes the proof of (240).

Finally, (241) is clear from interpretation of μ as the inverse of the function $x(t)$, since the discontinuity points, (233), of $x(t)$ are dense on the interval $0 \leqq t \leqq 1$.

62. It is readily seen from the identity mentioned after (232) that, if $k = 1,\ldots,2^n$, the total variation of the monotone function $x(t)$ on the open interval $(k - 1)/2^n < t < k/2^n$ is the same for all 2^n values of k and depends, therefore, only on n. On the other hand, it is clear from §61 that

(242) $\mu(x) = t \text{ if } x(t - 0) < x < x(t + 0),$

whether t is or is not a continuity point of the monotone function $x(t)$. These two facts obviously imply that, if the measure (240) is positive, then $\mu(x)$ is a constant multiple of the Euclidean measure of the common part, $S(\mu)E^x$, of the spectrum, $S(\mu)$, and of the half-line, E^x, consisting of those points of the x-line which are to the left of the given x. Since $\mu(\infty) = 1$, the constant of proportionality must, of course, be the reciprocal value of the measure (240). Accordingly,

(243) $\mu(x) = \operatorname{meas}(S(\mu)E^x)/\operatorname{meas} S(\mu) \text{ if } \operatorname{meas} S(\mu) \neq 0.$

It is easy to see that an identity corresponding to (243) holds in the case meas $S(\mu) = 0$, if the Euclidean measure occurring in (243) is replaced by the appropriate Hausdorff measure.

The content of (242) and of its extension just mentioned is that the distribution (217) is *linear on the set* $S(\mu)$; a set which is nowhere dense and perfect, by (241) and (226). A corollary is that

(244) μ is, under the assumption (236), absolutely continuous or purely singular according as the measure (240) is positive or 0.

In fact, it is clear from (223) that, whether (236) is or is not satisfied, μ is purely singular whenever its spectrum is of Euclidean measure 0 (cf. § 4). On the other hand, if the Euclidean measure of $S(\mu)$ is positive, and if (236) is assumed, then (243) is applicable. Since (243) implies that μ is absolutely continuous, the proof of (244) is complete.

The truth of the alternative expressed by (244) is by no means obvious. For instance, (244) has no analogue in the case characterized by the assumption that

$$(245) \qquad r_m \geqq \sum_{n=m+1}^{\infty} r_n \text{ for every } m, \qquad (q_0 = \sum r_n < \infty);$$

an assumption which, in view of (235), represents the precise opposite of the situation represented by (236).

In fact, examples, to be given in § 63 below, show that, if (245) is satisfied, μ can be either absolutely continuous or purely singular.

On the other hand, it is easy to see that

(246) $S(\mu)$ is the interval $-q_0 \leqq x \leqq q_0$ in the case (245).

Since $q_0 = \sum r_n$, the assertion of (246) is that every number s satisfying $-\sum r_n \leqq s < \sum r_n$ admits a representation of the form (221). But this is clear from the assumption (245).

63. In order to illustrate the above results, let the positive numbers r_n be so chosen as to form a geometric progression, $r_n = r^n$, and let the resulting distribution, (217), be denoted by μ_r. Thus, from (218)

$$(247) \qquad \mu_r^{\circ}(y) = \prod \cos(r^n y).$$

The necessary and sufficient condition (219) for the existence of the distribution μ_r is $0 < r < 1$. Consequently, (222) must be satisfied.

Since (235) gives $q_m = q_0 r^m$, where $q_0 = \sum r^n = (r^{-1} - 1)^{-1}$, it is clear that (236) is satisfied by $r_n = r^n$ if and only if $r < \frac{1}{2}$; in which case the measure (240) is seen to be 0. On the other hand, (245) is satisfied by $r_n = r^n$ whenever $r \geqq \frac{1}{2}$. It follows therefore from (241), (226) and (246), that

(248) $S(\mu_r)$ is a bounded perfect set of Euclidean measure 0 or it is the interval $-(r^{-1} - 1)^{-1} \leqq x \leqq (r^{-1} - 1)^{-1}$ according as $0 < r < \frac{1}{2}$ or $\frac{1}{2} \leqq r < 1$.

If k is a positive integer and if $0 < r < 1$, then $0 < r^{1/k} < 1$ and, according to (247),

$$(249) \qquad \mu_{r^{1/k}}^{\circ}(y) = \mu_r^{\circ}(y)\, \mu_r^{\circ}(r^{1/k}y) \ldots \mu_r^{\circ}(r^{(k-1)/k}y).$$

In view of (17) and (53), this can be written in the form

$$(250) \quad \mu_{r^{1/k}}(x) = \mu_r(x) * \mu_r(r^{-1/k}x) * \ldots * \mu_r(r^{-(k-1)/k}x).$$

According to a classical identity, the infinite product (247) belonging to $r = \frac{1}{2}$ is the elementary entire function $(\sin y)/y$ (Vieta); so that

$$(251) \qquad \mu_{1/2}^{\circ}(y) = y^{-1} \sin y.$$

Hence it is seen from (10) that $\mu_{1/2}(x)$ is the absolutely continuous distribution for which the density is

$$(252) \qquad \mu_{1/2}'(x) = \tfrac{1}{2} \text{ if } |x| < 1 \text{ and } \mu_{1/2}'(x) = 0 \text{ if } |x| > 1.$$

It follows therefore from (76) and from the case $r = 1/2$ of (250), that

$$(253) \quad \mu_r \text{ is absolutely continuous if } r = 2^{-1/k}, \qquad (k = 1, 2, \ldots).$$

It is clear from (223) and (248) that

$$(254) \qquad \mu_r \text{ is purely singular if } 0 < r < \tfrac{1}{2}.$$

Since application of (250) to $r = 1/4$, $k = 2$ gives

$$(255) \qquad \mu_{1/2}(x) = \mu_{1/4}(x) * \mu_{1/4}(\tfrac{1}{2}x),$$

and since (254) implies that the first, hence also the second, distribution on the right of (255) is purely singular, the truth of the assertion (77) follows from the case $k = 1$ of (253).

It is easy to show that, in contrast to (251),

$$(256) \qquad \limsup_{y \to \infty} \mu_{1/m}^{\circ}(y) > 0 \text{ if } m = 3, 4, \ldots \; .$$

In fact, it is seen from (247) that

$$\mu_r^{\,0}(2\pi/r^j) = \prod \cos\,(2\pi r^n) \text{ if } r = 1/m, \tag{257}$$

whenever m (> 1) and j are positive integers. Since $2\pi/r^j \longrightarrow \infty$ as $j \longrightarrow \infty$, and since the product on the right of (257) is independent of j and has, unless m = 2, a positive value, the assertion (256) follows from the identity (257).

This situation can be imitated in the case in which 1/r, instead of being a rational integer m, is an algebraic integer of sufficiently strong rational affinity. Such an algebraic integer is, for instance, the quadratic irrationality

$$r = \tfrac{1}{2}(5^{\frac{1}{2}} - 1); \text{ so that } \tfrac{1}{2} < r < 1, \tag{258}$$

hence $1/r = \frac{1}{2}(5^{\frac{1}{2}} + 1) > 1$. Here $s = 1/r$ and $-r$ are the roots of the equation $s^2 - s - 1 = 0$. More generally, let r, where $0 < r < 1$, be an irrational number such that $s = 1/r$ is an algebraic integer having the property that, if $s,t,\ldots,z$ denotes the set of all algebraic conjugates of s, then $|r| < 1, \ldots, |z| < 1$ (while $s > 1$, since $s = 1/r$). Since $s^n + t^n + \ldots + z^n$ is a real, rational integer for $n = 1,2,\ldots$, there exists on the interval $0 < \theta < 1$ a number θ such that the absolute value of the difference between s^n and the real, rational integer closest to s^n does not exceed a fixed positive multiple of θ^n as $n \longrightarrow \infty$. Hence there exists a positive constant c such that

$$\prod_{n=0}^{j-1} |\cos\,(\pi s^n)| > c \prod \cos\,(\pi\theta^n) \tag{259}$$

for every positive integer j. Consequently, if θ, where $0 < \theta < 1$, is so chosen that $\cos\,(\pi\theta^n) \neq 0$ for every n, i.e., that $\theta \neq \frac{1}{2}$, then the finite product on the left of (259) is not less than the positive constant ca, where a denotes the value of the infinite product on the right of (259). On the other hand, since $r = 1/s$, it is seen from (247) that, corresponding to (257),

$$(260) \qquad \mu_r^{\,o}(\pi/r^j) = \mu_r^{\,o}(\pi) \prod_{n=0}^{j-1} \cos(\pi s^n),$$

where $j = 1,2,\ldots$. Finally, since r is an algebraic, and π a transcendental, number, none of the factors of (247) vanishes when $y = \pi$, and so $\mu_r^{\,o}(\pi) \neq 0$. Since the product on the left of (259) was minorized by the positive constant ca for every j, and since $\mu_r^{\,o}(\pi) \neq 0$, it follows from (260) that

$$(261) \qquad \limsup_{y \to \infty} |\mu_r^{\,o}(y)| > 0;$$

in fact, $\pi/r^j \to \infty$ as $j \to \infty$, since $0 < r < 1$.

According to (191) and (223), the inequality (261) implies that the distribution μ_r is purely singular. This is not contained in (254), since some of the r-values satisfying (261) are outside the range $0 < r < \frac{1}{2}$, as shown by the example (258).

For the same reason, it now follows from (248) that there exist r-values for which the distribution μ_r is purely singular but such that the spectrum $S(\mu_r)$ is an interval (the existence of such distributions was observed in §4). The truth of the negation (190) follows not only from these examples but from the simpler result (256) as well.

64. It turns out that the r-values of the range $0 < r < 1$ which satisfy (261) are quite exceptional. In fact, it will be shown that (261) cannot hold unless r is so chosen as to fail to possess the following property: There does not exist any positive number b such that, if d_n denotes the absolute value of the difference between b/r^n and the integer closest to b/r^n, then

$$(262) \qquad d_n < \tfrac{1}{2}(1 + 1/r)^{-2} \text{ as } n \longrightarrow \infty,$$

In view of (18), the assertion can be formulated as follows: If

$$(263) \qquad \lim_{y \to \pm\infty} |\mu_r^{\,o}(y)| = 0$$

does not hold, then there must exist to r some b satisfying (262). But it is known that the set of those values in the interval $0 < r < 1$ to which there exists at least one $b = b(r)$ satisfying (262) is enumerable (Pisot).

It is trivial that, if $1/r$ is rational but not an integer, then there cannot exist to r any b satisfying (262). Hence, what will be shown implies that (263) holds for every rational r not exempted by (256).

Let r, where $0 < r < 1$, be any value for which there does not exist any b satisfying (262). Suppose, if possible, that the assertion, namely (263), is false for this r. Then there exist a sequence $y_1, y_2, \ldots$ and a positive bound $c = c(r)$ such that $|\mu_r^{\,o}(y_j)| > c$ and $y_j \longrightarrow \infty$ as $j \longrightarrow \infty$. In view of (247), the sequence $y_1, y_2, \ldots$ can be assumed to be so chosen that, if k_j denotes the (unique) positive integer for which $r < y_j r^{k_j} \leqq 1$, then the sequence $y_1 r^{k_1}, y_2 r^{k_2}, \ldots$ tends to a limit. Let this limit be denoted by u; so that $u_j \to u$, if $u_j = y_j r^{k_j}$. Since $y_j \longrightarrow \infty$, it is clear that $k_j \rightarrow \infty$. But (247) shows that, since $u_j = y_j r^{k_j}$,

$$(264) \qquad \mu_r^{\,o}(y_j) = \prod \cos\,(u_j r^{n-k_j}),$$

where j is arbitrarily fixed. Since $|\mu_r^{\,o}(y_j)| > c > 0$, the absolute value of the infinite product on the right of (264) exceeds a fixed positive constant. Hence, if this product is treated in the same way as the product (260) was treated, it is seen by a straightforward adaptation of the proof of (261) in § 63, that the limit relations $u_j \rightarrow u$ and $k_j \rightarrow \infty$ contradict the assumption, according to which r is such that there does not exist any b satisfying (262).

This proves, in particular, that the r-values satisfying (263) are dense on the interval $0 < r < 1$, and therefore on the interval $0 < r < \frac{1}{2}$. Hence, the truth of the negation (191) follows from (254).

On the other hand, the algebraic r-values of the type (258) and the rational r-values specified in (256) do not satisfy (263). Furthermore, it is clear from (249) that, if (263) does not hold for a given r, say for $r = r_0$, then (263) cannot hold for any r of the form $r = r_0^{\,a}$, where a is any rational exponent greater than 1 (so that $r_0^{\,a} < r_0$). It is not known whether the r-values satisfying (261) do or do not cluster at the upper bound, 1, of the range $0 < r < 1$.

The above examples imply that the decision of the alternative represented by (261) and (263) depends on delicate arithmetical properties of r. It is therefore of interest that, if $\mu_r^{\,o}(y)$ is replaced, as in (224), by its square-average, then all the Diophantine intricacies disappear. In fact, it will now be shown that

$$(265)\qquad \frac{1}{t}\int_0^t |\mu_r{}^{\circ}(y)|^2\, dy = O(t^{1/\log_2 r}) \text{ as } t \longrightarrow \infty$$

holds for every $r \leq \frac{1}{2}$. Since the logarithm in (265) is of base 2, the assertion (265) is trivial from (224) if $r = \frac{1}{2}$. The remaining case, where the exponent on the right of (265) is negative, is precisely the case in which $S(\mu_r)$ is of Euclidean measure 0; cf. (248). And it is readily seen that the Hausdorff *dimension of* $S(\mu_r)$, referred to after (243), is exactly the one corresponding to the exponent $\log_2 r$ in the mean estimate, (265), of the *transform of* μ_r. In view of (251) and (252), this parallelism holds in the limiting case $r = \frac{1}{2}$ also. Correspondingly, it is easy to show that the exponent on the right of (265) cannot be improved for any $r < \frac{1}{2}$.

In this connection, it is worth observing that both the proof and the assertion of (148) remain valid (without any change), if the notions of an absolutely continuous and of a purely singular function are referred to *any* Hausdorff measure, rather than, as in § 4, the *specific* Euclidean measure.

In order to prove (265), choose a positive integer k and omit all but the first k factors of the infinite product (247). Then, since $|\cos^2 u| < 1$ and $\cos u = \frac{1}{2}(e^{iu} + e^{-iu})$,

$$\int_0^t |\mu_r{}^{\circ}(y)|^2\, dy < 2^{-2k}\int_0^t \Big|\sum_{\pm} \exp i(\pm r \pm r^2 \pm \dots \pm r^k)y\Big|^2\, dy,$$

where the summation refers to all 2^k variations of the k alternatives $\pm, \dots, \pm$. If the integrand, on the right, $|\quad|^2$, is written as a product of complex conjugates, then it appears as a linear combination of terms of the form exp (icy). The various values of c, along with the number of terms belonging to a fixed value of c, can be obtained by an easy counting. Since the assumption $0 < r < \frac{1}{2}$ implies that the sum of the geometric series $r^{m+1} + r^{m+2} + \dots$ is less than r^m for every non-negative integer m, it then follows from the last formula line, that

$$(266)\qquad \int_0^t |\mu_r^o(y)|^2\, dy < 2^{-2k}\,\text{const.}(2^k t + 2^k \sum_{j=0}^{k} 2^{k-j} r^{-j}),$$

where the constant depends only on r. In fact, if exp(icy) is integrated between y = 0 and y = t, the corresponding term acquires the factor t or the divisor c according as c = 0 or c ≠ 0. Since 2r < 1, it is clear from (266) that

$$\int_0^t |\mu_r^o(y)|^2 dy < 2^{-k}\,\text{Const.}(t + 1/r^k),$$

where the constant depends only on r. Now let k be made a function of t, namely the integral part of the quotient (log t)/(-log r). Then the last inequality becomes identical with (265), where $0 < r < \frac{1}{2}$.

65. None of the above distributions is absolutely continuous but such as to have a nowhere dense spectrum. Such an example results by choosing

$$(267)\qquad r_n = 2^{-n} + 3^{-n}$$

in (217). In fact, it is seen from (235) that (236) is satisfied, and that the limit (240) is distinct from 0, in the case (267). Hence the assertion follows from (244).

Still smoother distributions can be obtained from the identities (249) and (251). In fact, these identities imply that

(268) $$\mu_r^{o}(y) = O(|y|^{-k}) \text{ if } r = 2^{-1/k},$$

where $k = 1,2,\ldots$. It follows therefore from the criterion (197), that the distribution μ_r has at least $k - 2$ continuous derivatives for every x if $r = 2^{-1/k}$. These r-values tend to 1 as $k \longrightarrow \infty$. But $r = 1$ is excluded by (219), and it is not difficult to prove that none of the distributions μ_r has derivatives of arbitrarily high order for every x.

In order to obtain distributions (217) possessing this, or even a higher, degree of smoothness, denote by λ_a the distribution which results if r_n in (217) is chosen to be $1/n^a$; so that, from (218),

(269) $$\lambda_a^{o}(y) = \prod \cos(y/n^a), \text{ where } a > \tfrac{1}{2},$$

by (219). It will be shown that each of these distributions $\lambda_a(x)$ has derivatives of arbitrarily high order for $-\infty < x < \infty$; that λ_a is regular-analytic and bounded in a strip containing the x-axis, if $\frac{1}{2} < a$ is replaced by $\frac{1}{2} < a \leqq 1$; and that λ_a is a (transcendental) entire function, if $\frac{1}{2} < a < 1$.

If $1 < a$, then λ_a cannot be regular-analytic at every point of the line $-\infty < x < \infty$. In fact, if $1 < a$, then (226) shows that $S(\lambda_a)$ is a bounded set, since $r_n = 1/n^a$. But the derivative $\lambda_a'(x)$ vanishes at every x not contained in the spectrum, and so $\lambda_a'(x)$ cannot be regular-analytic at the points $x = \pm \sum 1/n^a$, for instance; points which, according to (221), are contained in the spectrum. Since the spectrum is situated between these two points, it also follows that all derivatives of $\lambda_a(x)$ must vanish at these extreme points, although $\lambda_a(x)$ is not constant there.

In order to prove that λ_a possesses the smoothness properties stated before, it will first be shown that there exists an absolute constant C satisfying

$$(270)\quad |\mu^o(y)| = \prod |\cos(r_n y)| < \exp(-Cy^2 \sum_{r_n<1/|y|} r_n^2)$$

for every y, where $\mu^o(y)$ is the transform, (218), of an arbitrary distribution (217). The existence of the function Σ on the right of (270) is assured by (219).

Since $|\cos(r_n y)| \leqq 1$, the absolute value of the infinite product (270) is not increased if some of the factors are omitted. For every fixed positive y, omit all those factors in which r_n is not less than $1/y$. Then, if a positive constant B is chosen so small that $0 < \cos t < 1 - Bt^2$ for $0 < t < 1$, it follows that

$$|\mu^o(\pm y)| < \prod_{r_n<1/y} (1 - Br_n^2 y^2) = \exp \sum_{r_n<1/y} \log(1 - Br_n^2 y^2)$$

holds for every $y > 0$. This proves the truth of the inequality for a certain positive constant $C = C_B$, since $\log(1 - t) < -t$ for $0 < t < 1$.

If $\mu = \lambda_a$, where $a > \frac{1}{2}$, then the sum multiplying $-Cy^2$ on the right of (270) is

$$\sum_{1/n^a<1/|y|} (1/n^a)^2 = \sum_{|y|<n^a} n^{-2a} \sim \int_{|y|^{1/a}}^{\infty} u^{-2a}\, du$$

as $y \longrightarrow \pm\infty$. Since the last integral is $\text{const.}|y|^{-2+1/a}$, where const. = $(2a - 1)^{-1} > 0$, it follows from (270) that there exists to every $a > \frac{1}{2}$ a positive constant c_a satisfying

$$(271)\qquad |\lambda_a^o(y)| < \exp(-c_a|y|^{1/a}) \text{ for every } y.$$

Consequently, the smoothness properties stated after (269) follow from the corresponding sufficient criteria, (198) and (200), of § 53.

According to (226), the spectrum of the distribution $\mu = \lambda_a$ is the whole line $-\infty < x < \infty$ exactly in the case, $a \leqq 1$, in which $\lambda_a(x)$ is regular-analytic along this line. However, the distribution (217) can be so chosen as to be purely singular and still such that its spectrum is the whole line. According to (226), (225) and (191), a distribution (217) will certainly be of this type if it belongs to a series (219) for which the series $\sum r_n$ diverges but, as $y \rightarrow \pm\infty$, the product (218) does not tend to 0. And these conditions are satisfied by the sequence $r_1, r_2, \ldots$ which consists of the values $1/2^{k!}$, where $k = 1,2,\ldots$, and contains the k-th of these values exactly $2^{k!}$ times. In fact, the series $\sum r_n^2$ and $\sum r_n$ then become $\sum 2^{k!}/(2^{k!})^2 = \sum 1/2^{k!} < \infty$ and $\sum 2^{k!}/2^{k!} = \sum 1 = \infty$ respectively, and so it is sufficient to show that the corresponding product (218), namely

$$\mu^{\circ}(y) = \prod \cos^{2^{k!}}(y/2^{k!}), \tag{272}$$

does not tend to 0 as $y \rightarrow \pm\infty$. But

$$|\mu^{\circ}(2^{m!}\pi)| = \prod_{k=m+1} |\cos^{2^{k!}}(2^{m!-k!}\pi)|.$$

And it is easily verified that the value of the last product does not tend to 0 as $m \rightarrow \infty$, the situation being about the same as in (260), (261).

66. The general fact expressed by (225) concerns itself with the local behavior of a distribution (217) belonging to an arbitrary series (219). As a corresponding general fact concerning the behavior of the distributions (217) in the large, it will now be shown that, for an arbitrary series (219), the distribution $\mu(x)$ must tend, as $x \to \pm\infty$ to $\mu(\pm\infty) = \frac{1}{2}(1 \pm 1)$ so rapidly that all of its moments are finite and, what is much more,

$$\int e^{px^2}\, d\mu(x) < \infty \text{ for every } p > 0. \tag{273}$$

First, the spectrum of the distribution β_r, introduced at the beginning of § 57, is a bounded set. Hence, the same holds, by (71), for the distribution μ^n defined by

$$\mu^n = \beta_{r_1} * \dots * \beta_{r_n}, \tag{274}$$

where n and the positive numbers $r_1, \dots, r_n$ are arbitrary. Thus it is clear from (30) that all moments of the distribution μ^n are finite. The numerical values of these moments can be calculated by observing that, since $\beta_r^{\,0}(y) = \cos(ry)$ for every y, the transform of the distribution (274) is the product $\cos(r_1 y) \dots \cos(r_n y)$, by (82). It then follows from (32) that the 2h-th derivative of this product at the point $y = 0$ is $(-1)^h (\mu^n)_{2h}$ for every non-negative integer h. Accordingly,

$$(\mu^n)_{2h} = (\sum \dots \sum) C_{2j\ 2k\ \dots\ 2m} r_1^{2j} r_2^{2k} \dots r_n^{2m}, \tag{275}$$

where the summation, $(\sum \dots \sum)$, extends over all sets $(j, k, \dots, m)$ of non-negative integers subject to the restriction $2j + 2k + \dots + 2m = 2h$, i.e., $j + k + \dots + m = h$. It is understood that the C denote the corresponding multinomial coefficients; so the coefficient $C_{2j\ 2k\ \dots\ 2m}$ results if in

(276) $C_{j\ k\ \dots\ m} = (j' + k + \dots + m)/(j!k!\dots m!)$, where $j + k + \dots + m = h$,

$j, k, \dots, m$ and h are replaced by $2j, 2k, \dots, 2m$ and $2h$ respectively.

It is readily seen that $C_{2j\ 2k\ \dots\ 2m} \leqq h^h C_{j\ k\ \dots\ m}$. Hence, (275) shows that $(\mu^n)_{2h}$ is not greater than h^h times a sum which, according to the multinomial theorem, is exactly the sum representing the h-th power of $r_1^2 + \dots + r_n^2$. Consequently, $(\mu^n)_{2h} < h^h A^h$, if A denotes the sum of the infinite series (219). Since the bound $h^h A^h$ is independent of n, and since $\mu^n \longrightarrow \mu$ as $n \longrightarrow \infty$ holds in view of the definitions (274) and (217), it is now clear from (5*) that $(\mu)_{2h} \leqq h^h A^h$.

This means that $\int x^{2h} d\mu(x) < (hA)^h$, where $h = 0, 1, \dots$. It follows therefore from (4*), by placing $t = px^2$ in $e^t = 1 + \sum t^h/h!$, that the value of the integral (273) belonging to a fixed $p > 0$ cannot exceed $1 + \sum (phA)^h/h!$ and must, therefore, be finite if the positive series $\sum (Bh)^h/h!$, where $B = pA$, is convergent. But Stirling's formula shows that this series is convergent if $B < 1/e$. Consequently, the integral (273) must be finite if $p < 1/(Ae)$.

In order to complete the proof of (273), it remains to be shown that the restriction $p < 1/(Ae)$ is superfluous. But A denotes the sum of the convergent series (219). Hence, $1/(Ae)$ can be made arbitrarily large if the sequence $r_1, r_2, \dots$ is replaced by $r_{n+1}, r_{n+2}, \dots$, where n is sufficiently large. Then the distribution (217) becomes replaced by the distribution $\mu_{n+1} * \mu_{n+2} * \dots$. If the latter is denoted by μ_n, then $\mu = \overset{n}{\mu} * \mu_n$, where μ^n is the distribution (274). But μ^n has a bounded spectrum for every n, and so the integral (273) belonging to μ^{n} is finite for every p. Since the integral (273) belonging to a given p and to μ_n was seen to be finite when n is sufficiently large, the truth of (273) for $\mu = {}^{n}\mu * \mu_n$ follows by observing that the relevant criterion of § 22 is applicable to $f(x) = e^{px^2}$ (for every fixed $p > 0$).

67. The above proof of (273) is to a large extent independent of the assumption (217). In fact, the proof remains valid without any change, if (217) is replaced by any convergent convolution $\mu = *\mu_n$ in which every μ_n has moments of arbitrarily high order and $(\mu_n)_{2h}$ is less than the h-th power of a fixed multiple of $(\mu_n)_2$ for all n and for all h, finally μ_n is symmetric for every n.

A distribution μ is called symmetric if it is identical with the distribution ${}_{-1}\mu$ defined before (17). For instance, an absolutely continuous μ is symmetric if and only if it has a density μ' satisfying $\mu'(x) = \mu'(-x)$.

It is clear (18) and (19) that a distribution μ is symmetric if and only if the transform $\mu^{o}(y)$ is an even and/or real-valued function of y In fact, (10) becomes

$$\mu^{o}(y) = \mu(+0) - \mu(-0) + 2 \int_{+0}^{\infty} \cos(yx)\, d\mu(x). \tag{277}$$

Correspondingly, (185) can be contracted into

$$\frac{\mu(x+0) + \mu(x-0)}{2} = C + \frac{2}{\pi} \int_{0}^{\infty} \frac{\sin(xy)}{y}\, \mu^{o}(y)\, dy. \tag{278}$$

and so (195), if (195) is applicable, into

$$(279) \qquad \mu'(x) = 2\pi^{-1} \int_0^\infty \cos(xy)\, \mu^0(y)\, dy.$$

Since μ is symmetric if and only if μ^0 is real, it is clear from (53) and (28) respectively, that the convolution of two symmetric distributions and the limit of a convergent sequence of symmetric distributions are symmetric distributions.

Finally, it is clear from (30) that a symmetric μ cannot possess a non-vanishing moment of odd order. However, this does not imply that μ must be symmetric if all of its moments of odd order exist and vanish; cf. § 18.

68. In applications of the type exemplified by the theory of errors, of particular interest is the case of those densities of probability which are monotone functions of the distance from the origin. In order to delimit a corresponding class of distributions in such a way that the class becomes closed under the limit process specified at the end of § 2, the following definition turns out to be appropriate:

Let a distribution μ be called convex if it is symmetric and such that the curve $\mu = \mu(x)$ in an (x,μ)-plane is convex from below on the open half-line $-\infty < x < 0$, and therefore convex from above on the open half-line $0 < x < \infty$. Since μ is monotone, hence L-measurable, it follows from the theory of convex functions that, if μ is a convex distribution, the function $\mu(x)$ is absolutely continuous on every open interval not containing the point $x = 0$; a point at which μ can have a jump, as exemplified by the convex distribution (15). It also follows from the theory of convex functions (Jensen) that μ has a finite right-hand and a finite left-hand derivative at every $x \neq 0$; that both of these derivatives are monotone non-increasing functions of $|x|$; finally, that both of these derivatives are continuous at exactly those points $x \neq 0$ at which they have a common value (which implies that, since both derivatives are monotone, the set of points x at which μ has no ordinary derivative is enumerable). Both derivatives, being monotone, must tend to 0 as $x \rightarrow \pm\infty$, since $\int d\mu(x) < \infty$. But the derivatives can tend to $\mp\infty$ as $x \rightarrow \pm 0$.

If a sequence of functions which are convex on an open interval tends to a limit at every point of this interval, then the limit function obviously is convex there. This implies that, if $\mu_n \rightarrow \mu$ holds in the sense defined at the end of § 2, and if every μ_n is a convex distribution, then μ is a convex distribution.

This is the closure property of the class of convex distributions, referred to above. It holds only in virtue of the above definition of the class, since μ can have a jump at $x = 0$ when every μ_n is continuous at $x = 0$. On the other hand, it is clear that there exists to every convex distribution μ a sequence of convex distributions μ_n such that every μ_n has continuous derivatives of arbitrarily high order for every x, and $\mu_n \rightarrow \mu$.

It will now be verified that, if two distributions, say μ and λ, are convex, then $\mu * \lambda$ is a convex distribution.

In view of (56), it is sufficient to prove this for the case where μ and λ have continuous second derivatives. Then, since μ and λ are convex, $\mu'(x)$, $\lambda'(x)$ are even functions which do not increase when $|x|$ increases, i.e., the second derivatives $\mu''(x)$, $\lambda''(x)$ are odd functions which are non-positive for $x \geqq 0$. On the other hand, (81) shows that $\mu * \lambda$ has the density $\int \mu'(x - u)\lambda'(u)\, du$, and therefore the second derivative $\int \mu''(x - u)\lambda'(u)du$. Since μ'' is odd and λ' is even, this integral, where $-\infty < u < \infty$, can be written in the form

$$\int_{-\infty}^{+\infty} \lambda'(x - u)\mu''(u)\, du = \int_{0}^{\infty} \mu''(u) \left\{\lambda'(x - u) - \lambda'(x + y)\right\} du,$$

if it is assumed that $x > 0$. But this implies that the function $\mu''(u)$ is non-positive, and that, since $|x - u| \leqq x + u$, the function $\{\ \}$ multiplying $\mu''(u)$ is non-negative, and so the integrand $\mu''(u) \{\ \}$ is non-positive, on the whole integration domain $0 \leqq u < \infty$. Accordingly, the function $\mu * \lambda$ of x has a non-positive second derivative for every $x \geqq 0$. Since $\mu * \lambda$, being a convolution of two symmetric distributions, is symmetric, it follows that $\mu * \lambda$ is a convex distribution.

As an illustration, consider the distributions μ_p defined by (202), where $0 < p \leqq 2$. If $p = 1$ and $p = 2$, then (204) and (205) show that μ is convex. In the remaining cases, μ_p was obtained, in § 56, by taking a fixed distribution λ_p, changing in it the unit of length on the x-axis, forming convolutions of the resulting distributions, and finally applying a limit process. Since all of these operations preserve convexity, it follows that μ_p is convex if λ_p can be chosen to be convex. But the distribution λ_p used in § 56 was the one having the density (215), which is a decreasing function of $|x|$. Consequently, all the distributions μ_p

defined by (202) are convex.

Since the density (252) defines a convex distribution, it follows by the above argument that, if $q_1, q_2, \ldots$ is any sequence of positive numbers for which the infinite convolution $*\mu_n$, where $\mu_n(x) = \mu_{1/2}(q_n x)$, is convergent, then the distribution $*\mu_n$ is convex. If $q_n = (n - \frac{1}{2})^{-1}$, it is easily verified from (251), (17) and from the canonical products of the even entire functions cos z, z^{-1} sin z, that the function $\prod \mu_n^{\circ}(y)$, i.e., the transform of the distribution $*\mu_n$, is identical with the function $\prod \cos(y/n)$. Consequently, the distribution λ_1, defined by the case a = 1 of (269), is convex.

Projections

69. At the beginning of § 34, there was introduced a "space" T of "points" t. It was assumed that there is given a field of sets F of points t such that every t and T itself are in the field. Finally, it was assumed that there is assigned to every set F in the field a value $\phi(F)$ in such a way that

$$(280_1)\quad \phi(F) \geqq 0; \quad (280_2)\quad \sum \phi(F_n) = \phi(\sum F_n); \quad (280_3)\quad \phi(T) = 1,$$

where $F_1, F_2, \ldots$ is any sequence of mutually disjoint sets F.

In the sequel, T will be chosen to be the product space $X_1 \cdot \ldots \cdot X_k$, where k is a fixed positive integer and $X_1, \ldots, X_k$ denote the lines $-\infty < x_1 < \infty$, ..., $-\infty < x_k < \infty$. In other words, T will be the k-dimensional Euclidean space, and t any point $(x_1, \ldots, x_k)$ of this space. The field of the sets F will be chosen to be the Borel field on T. Finally, any function ϕ satisfying the three conditions (280_1), (280_2), (280_3) will be called a k-dimensional distribution, or simply a distribution (on T).

If k = 1, all of this reduces to the situation introduced in § 1. However, unless the contrary is stated or implied, it will be assumed that the fixed value of k is at least 2.

The spectrum $S(\phi)$, the point spectrum $P(\phi)$ etc. can be defined in exactly the same way as in § 4, with the understanding that "interval" now means "k-dimensional interval" (i.e., a set of all points $(x_1, \ldots, x_k)$ such that x_j is between a_j and b_j, where $a_j < b_j$ and $j = 1, \ldots, k$). Correspondingly, the absolute continuity of ϕ means the existence of a (non-negative) Baire function $f = f(x_1, \ldots, x_k)$ having the property that, for every Borel set F, the value of $\phi(F)$ is identical with the integral of $f \, dx_1 \ldots dx_k$ over F; in which case f is called the density of ϕ.

The symbol $\phi_n \longrightarrow \phi$ should be defined in exactly the same way as $\mu_n \longrightarrow \mu$ is defined at the end of §2. The notion of the convolution of two distributions, a notion depending on the existence of the translation group, can be transferred from k = 1 to the case of an arbitrary dimension number k. Correspondingly, the definition (10) of the transform of a distribution becomes

$$(281) \qquad \phi^o(y_1,\dots,y_k) = \int\dots\int \exp i(x_1y_1 + \dots x_ky_k)d\phi,$$

where $(y_1,\dots,y_k)$ is any point of a k-dimensional Euclidean space; it being understood that the integral (281) is a Borel-Lebesgue integral with reference to the measure ϕ on the $(x_1,\dots,x_k)$-space, and that the integration is extended over the whole of this space. A glance at §7 - §10 shows that the proofs of the theorems of concerning the uniqueness, compactness, and continuity of the transform remain valid for any k. And the whole theory of infinite convolutions, including that of the semi-groups, can be copied word by word.

The moments of the first and the second order are of course defined by

$$(282) \qquad (\phi)_j = \int\dots\int x_j d\phi \text{ and } (\phi)_{hj} = \int\dots\int x_h x_j d\phi, \quad (j,h = 1,\dots k),$$

if $\int\dots\int (|x_1|+\dots+|x_k|)\, d\phi < \infty$ and $\int\dots\int(x_1^2 + \dots + x_k^2)\, d\phi < \infty$ respectively. In the first case (and therefore, in particular, in the second case), the point $(x_1,\dots,x_k) = ((\phi)_1,\dots,(\phi)_k)$ of the $(x_1,\dots,x_k)$-space is the "center" of the "mass distribution" represented by ϕ. And ϕ is called centered if its center is the point $(x_1,\dots,x_k) = (0,\dots,0)$. In the second case, it is seen from (280_3), (281), (282) and from Taylor's formula for functions of k variables, that

$$(283)\qquad \phi^{o}(y_1,\ldots,y_k) = 1 + i\left\{(\phi)_1 y_1 + \ldots (\phi_k) y_k\right\} - \tfrac{1}{2}Q_\phi(y_1,\ldots y_k) + o(y_1^2 + \ldots + y_k^2)$$

as $y_1^2 + \ldots + y_k^2 \to 0$, where Q_ϕ denotes the quadratic form

$$(284)\qquad Q_\phi(y_1,\ldots,y_k) = \sum_{h=1}^{k}\sum_{j=1}^{k} (\phi)_{hj}\, y_h y_j .$$

According to (282), this quadratic form is identical with $\int\ldots\int (x_1y_1 + \ldots + x_ky_k)^2 d\phi$. This obviously implies that the form (284) is non-negative definite and that its rank is identical with d, if the spectrum $S(\phi)$ is contained in a d-dimensional, but is not contained in a (d - 1)-dimensional, Euclidean space through the origin of the $(x_1,\ldots,x_k)$-space (it being understood that d = 0 corresponds to the "unit distribution", for which $S(\phi)$ consists of the origin). In particular, (284) is positive definite if and only if d = k.

70. The existence of the moments (282) will not be needed in the sequel; so that ϕ is an arbitrary distribution on the $(x_1,\ldots,x_k)$-space, where k has a given value greater than 1.

Let $X: -\infty < x < \infty$ be an oriented line through the origin of the $(x_1,\ldots,x_k)$-space. For a given x, let F_X^x denote the k-dimensional, open half-space consisting of those points $t = (x_1,\ldots,x_k)$ for which the x-coordinate of the point t_X is less than the given x, where t_X denotes the orthogonal projection of t on X. Then it is clear from (280_1), (280_2),

(280_3) that the function $\mu(x) = \phi(F_X^{\,x})$ satisfies the three conditions (8) characterizing a distribution on X. This one-dimensional distribution will be called the projection of the k-dimensional distribution ϕ on X, and it will be noted by $\phi_X = \phi_X(x)$.

Thus $\overset{o}{\phi_X}(y) = \int e^{ixy} d\phi(F_X^{\,x})$, by (10). Hence it is readily verified from the definition (281) and from Fubini's theorem, that

$$(285) \qquad \overset{o}{\phi_X}(y) = \overset{o}{\phi}(\alpha_1 y,\dots,\alpha_k y); \qquad -\infty < y < \infty ,$$

where $\alpha_1,\dots,\alpha_k$ denote the direction cosines of the oriented line X.

It is clear from (285) that

$$(286) \qquad \overset{o}{\phi_{X_1}}(y) = \overset{o}{\phi}(y,0,\dots,0),\dots, \overset{o}{\phi_{X_k}}(y) = \overset{o}{\phi}(0,\dots,0,y),$$

where $X_1,\dots,X_k$ denote the coordinate axes in the $(x_1,\dots,x_k)$-space. Needless to say, the k functions of the single variable y do not determine the function (281) of k independent variables. In other words, distinct k-dimensional distributions $\phi = \phi(F)$ can have one and the same set of projections

$$(287) \qquad \phi_{X_1}(x),\ \dots,\ \phi_{X_k}(x)$$

on the respective coordinate axes $X_1,\dots,X_k$.

(On the other hand, if the projection $\phi_X(x)$ of ϕ is given for every X, i.e., if the continuum of all functions (285) belonging to a fixed function (281) is known, then the latter, and therefore the k-dimensional distribution ϕ, is uniquely determined. But not even this seems to be trivial. Its truth can be concluded from the fact that the solution of a certain integral equation of Abel's type is unique.)

71. Let R denote a rotation of the $(x_1,\ldots,x_k)$-space about the origin, and let RF be the Borel set into which a Borel F is turned by R. If a distribution ϕ has the property that $\phi(F) = \phi(RF)$ holds for every F and for every R, the distribution ϕ will be called of radial symmetry.

Since the scalar product $x_1y_1 + \ldots + x_ky_k$ remains invariant when the $(y_1,\ldots,y_k)$-space is subjected to the rotation R^{-1} inverse to the rotation of the $(x_1,\ldots,x_k)$, and since every distribution ϕ is uniquely determined by its transform (281), it is clear that ϕ is of radial symmetry if and only if the function (281) can be expressed as a function of the single variable

$$q = (y_1^2 + \ldots + y_k^2)^{\frac{1}{2}}, \text{ where } (\quad)^{\frac{1}{2}} \geqq 0. \tag{288}$$

On the other hand, the projection $\phi_X = \phi_X(x)$ of ϕ on the line X becomes independent of the choice of the line X, and can therefore be denoted simply by $\mu = \mu_\phi$. Correspondingly, the transform of $\phi_X = \mu_\phi$ becomes independent of the direction cosines $\alpha_1,\ldots,\alpha_k$ of X. Consequently, from (285) and (288),

$$\overset{o}{\mu_\phi}(q) = \phi^o(y_1,\ldots,y_k). \tag{289}$$

If a point $t = (x_1,\ldots,x_k)$, where $k > 1$, is thought of as a Borel set, and if ϕ is of radial symmetry, the three conditions (280_1), (280_2), (280_3) obviously imply that $\phi(t) = 0$ unless $t = (0,\ldots,0)$. In other words, the point spectrum, $P(\phi)$ is either vacuous or consists of the origin. Thus, if $\phi(0)$ denotes the value of $\phi(F)$ when F is the point $(0,\ldots,0)$, the distribution ϕ is or is not continuous according as $\phi(0) = 0$ or $\phi(0) > 0$. In either case, it is clear from the definition of the projection $\mu_\phi = \mu_\phi(x)$, that

$$\mu_\phi(+0) - \mu_\phi(-0) = \phi(0). \tag{290}$$

But it is also clear that the distribution μ_ϕ is symmetric in the sense defined in § 67. Consequently, from (277),

$$\overset{0}{\mu}_\phi(q) = \phi(0) + 2\int_{+0}^{\infty} \cos(qr)\, d\mu_\phi(r). \tag{291}$$

Let β denote that distribution of radial symmetry on the $(x_1,\ldots,x_k)$-space for which the spectrum $S(\phi)$ is the sphere $x_1^2 + \ldots + x_k^2 = 1$. Since the spectrum $S(\beta)$, being a $(k - 1)$-dimensional manifold, is of Euclidean measure 0, and since the point spectrum $P(\beta)$ is vacuous, the distribution β is purely singular. The distribution β will be called the equidistribution on the sphere $x_1^2 + \ldots + x_k^2 = 1$. The equidistribution on the sphere $x_1^2 + \ldots + x_k^2 = r^2$ of radius r is defined similarly; it will be denoted by β_r. (This notation agrees with the one used at the beginning of § 57 in the one-dimensional case; a case which is now excluded.) Thus, from (291),

$$\overset{0}{\mu}_{\beta_r}(q) = \overset{0}{\mu}_\beta(rq), \tag{292}$$

where $\beta = \beta_1$ and

$$\overset{0}{\mu}_\beta(q) = 2\int_{+0}^{\infty} \cos(qr)\, d\mu_\beta(r). \tag{293}$$

For an arbitrary distribution ϕ of radial symmetry, let $\lambda_\phi(r)$ denote the value of $\phi(F)$ when F is the interior of the sphere $S(\beta_r)$, i.e., the domain $x_1^2 + \ldots + x_k^2 < r^2$. Thus $\lambda_\phi(r)$ is a non-decreasing function satisfying

$$\lambda_\phi(+0) = \phi(0) \text{ and } \lambda_\phi(\infty) = 1; \tag{294}$$

cf. (290) and (280_3). If the distribution ϕ is thought of as a stratification of the possible jump (290) and of the spherical equidistributions, it is seen from (289) and (281) that, in view of Fubini's theorem,

$$(295) \qquad \mu_\phi^{\,o}(q) = \phi(0) + \int_{+0}^{\infty} \mu_{\beta_r}^{\,o}(q)\, d\lambda_\phi(r).$$

72. The relations (291) and (295) are two geometrical decompositions of the transform (281), represented by (289) and (288). In fact, the distribution $\mu_\phi(x)$ was defined in § 71 as the projection of ϕ on a line $-\infty < x < \infty$ through $(x_1,\ldots,x_k) = (0,\ldots,0)$; so that $\mu_\phi(x)$ is a symmetric distribution which is continuous except for a possible jump, (290), at the origin, and so $\mu_\phi(r) = 1 - \mu_\phi(-r)$ holds for every $r > 0$ (cf. the definition of ${}_{-1}\mu$ in § 6). Accordingly, the difference $\mu_\phi(r) - \mu_\phi(-r)$ is, on the one hand, identical with $2\mu_\phi(r) - 1$ and represents, on the other hand, the value of $\phi(F)$ for that Borel set F which is the k-dimensional strip contained between two parallel hyper planes tangent to the sphere of radius r about the origin. Hence, if the factor 2 in (291) is moved beneath the integral sign, it is seen from $2d\mu_\phi(r) = d\left\{2\mu_\phi(r) - 1\right\}$ that (291) represents the decomposition of the transform (289) into cylindrical waves. On the other hand, the definition of $\lambda_\phi(r)$ at the end of § 73 shows that (295) is the decomposition of the same transform into spherical

waves. Originally, the transform was given by the k-fold integral (281) in the form of plane waves.

Either of the weight functions, $2\mu_\phi(r) - 1$ and $\lambda_\phi(r)$, of the cylindrical and spherical stratifications determines the other uniquely. The explicit connection between these two functions is

$$(296) \qquad 2\mu_\phi(r) - 1 = \lambda_\phi(+0) + \int_{+0}^{\infty} \mu_\beta(r/p)\, d\lambda_\phi(p);$$

cf. (294). In fact, it is clear from (292) and (10) that $\mu_\beta(x/p)$ is the projection of the equidistribution carried by the sphere $x_1^2 + \dots + x_k^2 = p^2$, since $\mu_\beta(x)$ is the projection of the equidistribution carried by the sphere $x_1^2 + \dots + x_k^2 = 1$. Thus (296) is evident from the geometrical interpretation given before, the two evaluations being identical in view of Fubini's theorem. Correspondingly, (296) follows from the uniqueness result of § 7 also, if (296) is substituted into (291) and the resulting representation of $\overset{0}{\mu}_\phi(q)$ is compared with the one which results by substituting (292) and (293) into (295).

The function μ_β occurring in (296) can be given explicitly, as follows:

$$(297) \quad \mu_\beta(r) = \begin{cases} \frac{1}{2} + B_k \int_0^r (1 - x^2)^{\frac{1}{2}(k-3)}\, dx \text{ if } 0 \lesssim r \leq 1, \\ 1 \text{ if } 1 \leqq r < \infty, \text{ where } B_k = \pi^{-\frac{1}{2}}\, \Gamma(\tfrac{1}{2}k)/\Gamma(\tfrac{1}{2}k - 1), \end{cases}$$

k being the dimension number. In fact, $\beta = \beta(F)$ is the equidistribution on the sphere $x_1^2 + \dots + x_k^2 = 1$, and $\mu_\beta(x)$ is the one-dimensional distribution representing the projection of β on a line $-\infty < x < \infty$ through

$(x_1,\ldots,x_k) = (0,\ldots,0)$. Hence, $d\mu_\beta(x)$ is proportional to the $(k - 1)$-dimensional area of that portion of the sphere which is situated between the two hyper planes perpendicular to the x-line and proceed by the latter at the points x and $x + dx$. Thus $d\mu_\beta(x) = 0$ if either $-\infty < x < -1$ or $1 < x < \infty$, while $d\mu_\beta(x) = \text{const.} \sin^{k-1}\theta \, d\theta$ if $-1 < x < 1$, where θ denotes the azimuth in the $(x_1,\ldots,x_k)$-space with reference to the negatively oriented x-axis; so that $x = -\cos\theta$. Since this gives $d\mu_\beta(x) = \text{const.}(1 - x^2)^{\frac{1}{2}(k-3)}dx$ for $-1 < x < 1$, the assertion (297) follows, except for the numerical value of const. $= B_k$. But the latter is determined by (3); so that, corresponding to (215 bis),

$$\text{(297 bis)} \quad 1/B_k = 2\int_0^1 (1 - x^2)^{\frac{1}{2}(k-1)} dx = \pi^{\frac{1}{2}}\,\Gamma(\tfrac{1}{2}k - \tfrac{1}{2})/\Gamma(\tfrac{1}{2}k).$$

Substitution of (297) into (293) gives

$$\text{(298)} \qquad \overset{\circ}{\mu}_\beta(q) = 2B_k \int_0^1 (1 - r^2)^{\frac{1}{2}(k-1)} \cos(qr)\, dr.$$

This corresponds to (216) and can, corresponding to (216 bis), be expressed in terms of the cylinder function

$$\text{(298 bis)} \quad \frac{J_h(z)}{(\frac{1}{2}z)^h} = \frac{\pi^{-\frac{1}{2}}}{\Gamma(h + \frac{1}{2})}\, 2\int_0^1 (1 - s^2)^{h-\frac{1}{2}} \cos(zs) dx; \qquad h > -\tfrac{1}{2}.$$

In fact, $h = \frac{1}{2}k - 1$, and so, from (297 bis),

$$\text{(299)} \qquad \overset{\circ}{\mu}_\beta(q) = 2^{\frac{1}{2}k-1}\,\Gamma(\tfrac{1}{2}k)\, J_{\frac{1}{2}k-1}(q)/q^{\frac{1}{2}k-1}.$$

(Incidentally, the function on the right of (299) becomes $\cos q$ in the case $k = 1$ considered at the beginning of § 57; a case which is now excluded.) If $k = 3$, then (299) is reduced to the function (251), where $y = q$. If $k = 2$, then (299) becomes just $\overset{\circ}{\mu}_\beta = J_0$.

Since $J_h(q) = O(q^{-\frac{1}{2}})$ as $q \longrightarrow \infty$ holds for every $h > -\frac{1}{2}$, it follows from (299) that

$$(300) \qquad \overset{o}{\mu}_\beta(q) = O(q^{-\frac{1}{2}k+\frac{1}{2}}) \text{ as } q \longrightarrow \infty.$$

The estimate (300), when compared with (297), illustrates the conclusion drawn from (197) in § 53.

If (299) and (292) are substituted into (295), where ϕ is any k-dimensional distribution of radial symmetry, it follows that

$$(301) \qquad \overset{o}{\mu}_\phi(q) = \phi(0) + \Gamma(\tfrac{1}{2}k)(\tfrac{1}{2}q)^{1-\frac{1}{2}k} \int_{+0}^{\infty} r^{1-\frac{1}{2}k} J_{\frac{1}{2}k-1}(qr)\,d\lambda_\phi(r).$$

Accordingly, the k-fold integral (281) is reduced by (289) and (288) to the integral (301). Incidentally, (290), (291) and the parenthetical remark following (299) show that (301) is true in the excluded case $k = 1$ also, since then $\phi = \mu_\phi$ in virtue of § 5.

In view of (291) and (298 bis), the relation (301) can be interpreted as the harmonic analysis (Fourier transform) of the connection between the two functions of stratification, $\mu_\phi(r)$ and $\lambda_\phi(r)$, belonging to a distribution ϕ of radial symmetry. This connection, namely (296), can be written in a form as explicit as (301). In fact, $\mu_\beta(+0) = \frac{1}{2}$ and $\mu_\beta(\infty) = 1$, by (297). If this is combined with (294), a partial integration shows that the sum on the right of (296) is identical with

$$\tfrac{1}{2} - \int_{+0}^{\infty} \lambda_\phi(p)\,d_p\mu_\beta(r/p), \text{ i.e., } \tfrac{1}{2} - B_k \int_{r}^{\infty} \lambda_\phi(p)(1 - r^2/p^2)^{\frac{1}{2}(k-3)}(-r/p^2)\,dp,$$

by (297); so that (296) appears in the form

$$(302) \qquad \frac{4\mu_\phi(r) - 3}{2B_k r} = \int_{r}^{\infty} (p^2 - r^2)^{\frac{1}{2}(k-3)} \frac{\lambda_\phi(p)}{p^{k-1}}\,dp.$$

In view of (8), the numerator of the quotient on the left of (302) tends to 1 as $r \to \infty$.

In connection with (302), cf. (310) below.

If the function λ_ϕ is given, the function μ_ϕ follows from (302). In the converse problem, where μ_ϕ is given, (302) represents a functional equation for the unknown function λ_ϕ. It is clear that the solution of this functional equation can be reduced to a differentiation if $k = 3$, and to Abel's integral equation, that is, to a differentiation of order $\frac{1}{2}$, if $k = 2$. And it is easy to verify by successive partial integrations and by differentiations of integral order, that the solution of (302) in case of an arbitrary k can be reduced to the case $k = 3$ or to the case $k = 2$ according as k is odd or even. Correspondingly, the function (298 bis) occurring in the harmonic analysis (301) of (302) can, by successive differentiations of integral order, be reduced to the elementary function J_1 or to the less elementary function J_0 (resulting from J_1 by a differentiation of order $-\frac{1}{2}$), according as k is odd or even.

Needless to say, the integral equation (302) cannot have a non-negative. non-decreasing solution $\lambda_\phi(r)$ satisfying $\lambda_\phi(\infty) = 1$, unless the given function, $\mu_\phi(r)$, is such as to represent the projection of a k-dimensional distribution of radial symmetry.

73. Let $f(t) = f(x_1,\ldots,x_k)$ be a real-valued Baire function of the position on the k-dimensional Euclidean space $T = X_1 . \ldots . X_k$ on which there is given a distribution, that is, a set-function ϕ satisfying (280_1), (280_2), (280_3). The function f can be undefined or $\pm \infty$ at some points $(x_1,\ldots,x_k)$. However, the set, say F_f, of these points should be of ϕ-measure 0, i.e., such that $\phi(F_f) = 0$. Then § 34 assigns to f a ϕ-distribution, $\mu = \mu(x)$, as follows: $\mu(x)$ is the ϕ-measure of the set of those points $(x_1,\ldots,x_z)$ at which $f(x_1,\ldots,x_k)$ is less than x. (It is understood that the line X: $-\infty < x < \infty$ cannot be thought of as situated in the $(x_1,\ldots,x_k)$-space and must not, therefore, be confused with the line considered at the beginning of § 70).

Now let the given function f be

$$(303) \qquad f(x_1,\ldots,x_k) = \frac{\frac{1}{k} \sum_{j=1}^{k} x_j}{\left| \frac{1}{k} \sum_{i=1}^{k} (x_i - \frac{1}{k} \sum_{j=1}^{k} x_j)^2 \right|^{\frac{1}{2}}}$$

("Student's ratio"). The denominator, $| \quad |^{\frac{1}{2}}$, vanishes on a (k - 1)-dimensional manifold which represents the set F_f mentioned before. Accordingly, the distribution ϕ must be chosen such that this manifold, which is independent of ϕ, is of ϕ-measure 0. It is understood that k is assumed to be at least 2 (otherwise the denominator of (303) vanishes identically).

Suppose now that ϕ is of radial symmetry. It will be shown that the ϕ-distribution of the function (303) then is independent of the choice of ϕ. The explicit form of this common ϕ-distribution, a distribution which depends only on the dimension number k and can therefore be denoted by $\alpha_k = \alpha_k(x)$, will be a by-product of the proof. It turns out that α_k is the absolutely continuous distribution having a density proportional to $(1 + x^2)^{-\frac{1}{2}k}$ for $-\infty < x < \infty$; so that, from (215) and (215 bis),

$$(304) \quad \alpha_k'(x) = A_k(1 + x^2)^{-\frac{1}{2}k}, \text{ where } A_k = \pi^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}k + \tfrac{1}{2})/\Gamma(\tfrac{1}{2}k),$$

It must of course be assumed that the manifold, F_f, on which the denominator of (303) vanishes is of ϕ-measure 0. But ϕ is of radial symmetry, and so it is clear that $\phi(F_f) = 0$ if and only if $\phi(0) = 0$, where $\phi(0)$ denotes, as in § 71, the only possible jump of ϕ. Accordingly, the assertion is that the ϕ-distribution of (303) has the density (304) for every distribution ϕ which is of radial symmetry and such that $\phi(0) = 0$.

Let N denote the line which passes through the origin of the $(x_1,\dots,x_k)$-space and is normal to the hyper plane pierced by the x_j-axis at the point $x_j = 1$ of the latter, where $j = 1,\dots,k$. Then the k direction cosines of N have a common value, $k^{-\frac{1}{2}}$ (> 0, if N is oriented from the origin toward the hyper plane). Hence, if R_N denotes any rotation of the $(x_1,\dots,x_k)$-space about the line N, a rotation which is uniquely determined only when k = 2, then the k-rowed orthogonal matrix defining R_N has k equal elements, $k^{-\frac{1}{2}}$, in its k-th row. Accordingly, if $('x_1,\dots,'x_k)$ denotes the space into which the space $(x_1,\dots,x_k)$ is rotated by a given R_N, then

$$(305_1) \qquad 'x_k = \sum_{j=1}^{k} k^{-\frac{1}{2}} x_j, \text{ i.e., } \sum_{j=1}^{k} x_j = k^{\frac{1}{2}} \cdot {'x_k}.$$

Thus the square of the denominator of (303) is 1/k times

$$\sum_{i=1}^{k} (x_i - k^{-\frac{1}{2}}\,{}'x_k)^2 = \sum_{i=1}^{k} x_i^{\;2} - 2(k^{\frac{1}{2}}\,{}'x_k)\; k^{-\frac{1}{2}}\;{}'x_k + k\,(k^{-\frac{1}{2}}\;{}'x_k)^2.$$

But the first sum on the right is invariant under any rotation, and so the whole expression on the right is identical with the sum of the squares of the first k - 1 of the k coordinates $'x_j$. Accordingly, the denominator of (303) is

$$(305_2) \qquad \left|\quad\right|^{\frac{1}{2}} = (k^{-1} \sum_{j=1}^{k-1} {}'x_j^{\;2})^{\frac{1}{2}}.$$

Hence, if the numerator is expressed in terms of (305_1), it follows that the function (303) is transformed by any rotation R_N into the function

$$(306) \qquad {}'f({}'x_1,\ldots,{}'x_k) = {}'x_k/({}'x_1^{\;2} + \ldots + {}'x_{k-1}^{\;2})^{\frac{1}{2}}$$

(in the sense that $f(x_1,\ldots,x_k) = {}'f({}'x_1,\ldots,{}'x_k)$ in virtue of R_N). What will be essential in the proof of (304) is the fact that none of the k coordinates $'x_j$ occurs in the numerator and in the denominator of (306).

Let a set of points in the k-dimensional space be called a half-cone if it consists of a set of half-lines ending at the origin (so that "half-cone" means "half-cone issued from the origin"). Since the ratio (306) remains unchanged if each of the k coordinates $'x_j$ is multiplied by an arbitrary positive number, it is clear that, if C^x denotes the set of those points at which the value of the function (306) is less than a given number x, then C^x is a half-cone. It depends on the choice of the rotation R_N turning the original coordinate system, $(x_1,\ldots,x_k)$ into the coordinate system, $('x_1,\ldots,'x_k)$, which occurs in (306); so that $C^x = C^x(R_N)$. However, since ϕ is of radial symmetry, and since the function (306) is identical with the function (303) in virtue of R_N, the ϕ-measure of $C^x(R_N)$ is independent of R_N; in fact, it is identical with $\alpha_k(x)$, if $\alpha_k(x)$ denotes the ϕ-measure of the set of those points $(x_1,\ldots,x_k)$ at which the value of the function (303) is less than x.

Now let M be any (oriented) line through the origin. Let the coordinate system $(x_1,\ldots,x_k)$ be rotated into a coordinate system, say $(v_1,\ldots,v_k)$, which relates to the line M in the same way as the line N relates to the coordinate system $(x_1,\ldots,x_k)$. Finally, let $('v_1,\ldots,'v_k)$ denote the coordinate system belonging to the coordinate system $(v_1,\ldots,v_k)$ in the same way as $('x_1,\ldots,'x_k)$ belongs to $(x_1,\ldots,x_k)$. Then, since ϕ is of radial symmetry, the ϕ-measure of the set of those points $('u_1,\ldots,'u_k)$ at which the function $'u_k/('u_1^2 + \ldots + 'u_{k-1}^2)^{\frac{1}{2}}$, which corresponds to (306), is less than x, has the same ϕ-measure as before, namely $\alpha_k(x)$. Since this holds for any line M through $(0,\ldots,0)$, and since the group of all rotations about $(0,\ldots,0)$ is transitive, it now follows, by one more application of the radial symmetry of ϕ, that $\alpha_k(x)$ is identical with the ϕ-measure of the half-cone $C^x(R)$ for any choice of the rotation R which turns the coordinate system $(x_1,\ldots,x_k)$ into the coordinate system $('x_1,\ldots,'x_k)$ occurring in (306). Before this consideration, only rotations R of the particular type R_N were admitted.

Since $C^x(R)$ can be referred to an arbitrary R, it is now seen, by changing the notions in (306), that the function $\alpha_k(x)$, originally defined as the ϕ-distribution of (306), is the ϕ-distribution of

$$g(x_1,\dots,x_k) = x_k/(x_1^2 + \dots + x_{k-1}^2)^{\frac{1}{2}} \tag{307}$$

as well. In fact, all that this means is that $\alpha_k(x)$ is identical with the ϕ-measure of the half-cone on which the function (307) is less than x. Let this half-cone be denoted simply by C^x. Then, if the set RF defined as at the beginning of § 71, the radial symmetry of $\phi = \phi(F)$ implies that $\alpha_k(x) = \phi(RC^x)$ holds for every rotation R, since $\alpha_k(x) = \phi(C^x)$.

It is clear from (280_2) that $\alpha_k(x) = \phi(RC^x)$ cannot hold for every R unless $\alpha_k(x)$ is a constant multiple of meas c^x, where c^x denotes the set of those points of the sphere $x_1^2 + \dots + x_k^2 = 1$ which are on the half-lines constituting the half-cone C^x, and meas c^x is the $(k - 1)$-dimensional surface measure of c^x. But C^x was the set of those points $(x_1,\dots,x_k)$ of the k-dimensional space at which the function (307) is less than x. Hence, c^x is the set of those points of the sphere $x_1^2 + \dots + x_k^2 = 1$ at which the function $x_k/(1 - x_k^2)^{\frac{1}{2}}$ is less than x. Since the inequality $x_k/(1 - x_k^2)^{\frac{1}{2}} < x$ can be written in the form $x_k < s$, where

$$s = x/(1 + x^2)^{\frac{1}{2}}, \tag{308}$$

it follows that $\alpha_k(x)$ is a constant multiple of the $(k - 1)$-dimensional surface measure of that portion of the sphere $x_1^2 + \dots + x_k^2 = 1$ which is to the left of the hyper plane $x_k = s$, if the x_k-axis is thought of as oriented from left to right. But it was seen in the proof of (297) that this surface measure is a constant multiple of

$$\int_{-1}^{s} (1 - x_k^2)^{\frac{1}{2}(k-3)} \, dx_k, \tag{309}$$

if $-1 < s < 1$. Since (308) defines a one-to-onecorrespondence between the whole x-axis and the interval $-1 < s < 1$, it follows that $\alpha_k(x)$ is a constant multiple of (309), and therefore the derivative $\alpha_k'(x)$ is a constant multiple of $(1 - s^2)^{\frac{1}{2}(k-3)} s'$, for $-\infty < x < \infty$. This gives $\alpha_k'(x) = \text{const.}\ (1 + x^2)^{-\frac{1}{2}k}$, if $s = s(x)$ and $s' = s'(x)$ are substituted from (308). Since the numerical value of the factor of proportionality is determined by (3), the proof of (304) is complete.

74. The connecting link, (308), between the projections (297) of the spherical equidistributions and the distributions defined by (304) suggests for the latter distributions the following geometrical interpretation: Let the $(x_1, \dots, x_k)$-space be thought of as Riemann's spherical space on which $x_1, \dots, x_k$ are his Cartesian normal coordinates, mapping the "sphere" conformally on the Euclidean realization of the $(x_1, \dots, x_k)$-space; so that the coefficient matrix, (g_{ij}), of the squared line element becomes $(1 + x_1^2 + \dots + x_k^2)^{-2}$ times the unit matrix, if the unit of length is normalized the diameter of the "sphere". Since the square root of det (g_{ij}) is $(1 + r^2)^{-k}$, where $r^2 = x_1^2 + \dots + x_k^2$, the Riemannian volume element is $(1 + r^2)^{-k}$ times the Euclidean volume element $dx_1 \dots dx_k$. Consequently, if "equidistribution" is meant in the sense corresponding to the definition of the distribution β introduced after (291), then the equidistribution

on the Riemannian space is represented by that absolutely continuous distribution on the Euclidean $(x_1,\dots,x_k)$-space which has a density proportional to $(1 + r^2)^{-k}$. But the projection of this k-dimensional distribution on a line $-\infty < x < \infty$ through the origin $(0,\dots,0)$ is precisely the distribution $\alpha_k(x)$ found in § 73.

In order to verify this, let ϕ denote any absolutely continuous distribution of radial symmetry, and let $\delta_\phi(r)$, where $r = (x_1^2+\dots+x_k^2)^{\frac{1}{2}}$, denote the density of ϕ. According to the definition of the projection $\mu_\phi = \mu_\phi(x)$, introduced after (288), the value of the distribution μ_ϕ for a given x results if $\delta_\phi(r)\, dx_1 \dots dx_k$ is integrated over the half-space $x_k < x$; so that $d\mu_\phi(x)$, the value of $\phi = \phi(F)$ for the infinitesimal strip F contained between the hyper planes $x_k = x$ and $x_k = x + dx$, is dx times the integral of $\delta_\phi(r)\, dx_1\dots dx_{k-1}$ over the whole $(x_1,\dots,x_{k-1})$-space. Since $r = (x_1^2 + \dots + x_{k-1}^2 + x_k^2)^{\frac{1}{2}}$, it follows that the one-dimensional distribution $\mu_\phi(x)$ has the density

$$\mu_\phi'(x) = \int \dots \int \delta_\phi(\sqrt{p^2 + x^2}\,)\, dx_1\dots dx_{k-1}, \text{ where } p = (x_1^2 + \dots + x_{k-1}^2)^{\frac{1}{2}}.$$

But this p represents the radial polar coordinate in the (k - 1)-dimensional Euclidean space over which the integration is extended; so that $dx_1\dots dx_{k-1}$ is p^{k-2} times the surface element of the sphere $x_1^2 + \dots + x_{k-1}^2 = 1$ (if $k > 2$). Consequently,

$$(310) \qquad \mu_\phi'(x) = \text{const.} \int_0^\infty \delta_\phi(\sqrt{p^2 + x^2}\,)\, p^{k-2}\, dp,$$

where the const. is the (k - 1)-dimensional Euclidean surface area of the sphere (if $k > 2$; and const. = 2 in the limiting case $k = 2$). Incidentally, this surface area can be aalculated from (310) itself, if use is made of (315) below.

In the Riemannian case considered above, the density $\delta_\phi(r)$ was seen to be a constant multiple of $(1 + r^2)^{-k}$. Hence, the assertion to be verified is that (310) becomes identical with (304) in this case. But the truth of this becomes clear by observing that, if $\delta_\phi(r) = \text{Const.}(1 + r^2)^{-k}$, the replacement of p by the integration variable $q = p^2/(1 + x^2)$, where x is fixed, transforms the integral (310) into a constant multiple of $(1 + x^2)^{-\frac{1}{2}k}$.

Thus (304) and the derivative of (297) are the densities of the projections of the equidistributions carried by the k-dimensional spherical space and by the sphere $x_1^2 + \dots + x_k^2 = 1$ respectively. Consider now the equidistribution on the sphere $x_1^2 + \dots + x_k^2 < 1$. This is the k-dimensional distribution ϕ having a density $\delta_\phi(r)$ which is positive constant or 0 according as $r < 1$ or $r > 1$. Hence, the integral (310) becomes 0 or a constant multiple of the integral of $p^{k-1}dp$ over the interval $0 \leq p \leq (1 - x^2)^{\frac{1}{2}}$, according as $|x| \geq 1$ or $|x| \leq 1$.

Accordingly, the density of the projection $\mu_\phi(x)$ of the equidistribution ϕ carried by the sphere $x_1^2 + \dots + x_k^2 < 1$ is 0 or a constant multiple of $(1 - x^2)^{\frac{1}{2}(k-1)}$ according as $|x| > 1$ or $|x| < 1$. On the other hand, (297) states that the density of the projection $\mu_\beta(x)$ of the equidistribution β carried by the sphere $x_1^2 + \dots + x_k^2 = 1$ is 0 or a constant multiple of $(1 - x^2)^{\frac{1}{2}(k-3)}$ according as $|x| > 1$ or $|x| < 1$. Since the exponent, $\frac{1}{2}(k - 3)$, in the latter density, becomes identical with the exponent, $\frac{1}{2}(k - 1)$, in the former density if the dimension number of the $(x_1,\dots,x_k)$-space is increased by 2, it follows that the projection of the equidistribution carried by the sphere $x_1^2 + \dots + x_k^2 < 1$ is identical with the projection of the equidistribution carried by the sphere $x_1^2 + \dots + x_h^2 = 1$, where $h = k + 2$.

Actually, the identity of these two projections is just a restatement of a classical theorem in multi-dimensional stereometry. The limiting case $k = 1$, where the sphere $x_1^2 + \dots + x_k^2 < 1$ is the interval $-1 < x_1 < 1$, represents Archimedes' theorem, according to which the surface area of a sector cut off from the sphere $x_1^2 + x_2^2 + x_3^2 = 1$ by a plane $x_1 = x$ is proportional to the altitude $1 + x$ of the sector, where $-1 < x < 1$. In fact, if k is replaced by $h = k + 2$, then the function in the first line of (297) becomes linear in the case $\frac{1}{2}(h - 3) = 0$.

75. Instead of k-dimensional distribution ϕ of radial symmetry, consider for a moment the k-dimensional distributions ϕ which are product distributions. The latter are defined by the property that the ϕ-measure on the $(x_1,\dots,x_k)$-space is the product measure (in the sense of (119) and (120), where $t_j = x_j$, $n = k$ and $\phi^n = \phi$) of the projections of ϕ on the k coordinate axes X_j. It is clear from (118), (120) and from the uniqueness theorem of the transform (281), that a distribution ϕ is a product distribution if and only if

$$(311)\quad \phi^0(y_1,\dots,y_k) = \phi^0(y_1,0,\dots,0)\phi^0(0,y_2,\dots,0)\,\dots\,\phi^0(0,0,\dots,y_k)$$

holds for (281) as an identity on the $(y_1,\dots,y_k)$-space. In fact, the projections of ϕ on the coordinate axes and the transforms of these projections are given by (287) and (286) respectively; so that (311) is equivalent to

$$(312)\qquad \phi^{o}(y_1,\dots,y_k) = \phi_{X_1}^{\;o}(y_1)\,\phi_{X_2}^{\;o}(y_2)\,\dots\,\phi_{X_k}^{\;o}(y_k).$$

(Let $\phi_0(x)$ denote the one-dimensional distribution representing the projection of the k-dimensional product distribution ϕ on that line through the origin of the $(x_1,\dots,x_k)$-space which has the direction cosines $\alpha_j = k^{-\frac{1}{2}}$ for $j = 1,\dots,k$. Thus $\phi_0(x)$ is the ϕ-measure of the half-space $x_1 + \dots + x_k < x$; cf. (305_1). On the other hand, (285) and (312) show that

$$(312\text{ bis})\qquad \phi_0^{\;o}(k^{\frac{1}{2}}y) = \phi_{X_1}^{\;o}(y)\,\phi_{X_2}^{\;o}(y)\,\dots\,\phi_{X_k}^{\;o}(y).$$

This, when compared with (82), explains, via (123), the geometrical meaning of the convolution process.)

After this digression, consider again a question involving radial symmetry.

It is shown in the elements of the kinetic theory of gases that an absolutely continuous product distribution having a differentiable density cannot be of radial symmetry unless the density is a constant multiple $\exp(-r^2)$, if the unit of length is suitably chosen. It is easy to prove that this holds without any assumption of continuity, except that one has then to include the trivial case of the "unit distribution" (for which case the spectrum consists of the origin).

In fact, suppose that ϕ is a product distribution of radial symmetry. This means that, if q is defined by (288), both (289) and (311) are valid. Hence

(313) $\mu_\phi^{o}(q) = \mu_\phi^{o}(y_1)\,\mu_\phi^{o}(y_2)\,\dots\,\mu_\phi^{o}(y_k)$, where $q = (y_1^2 + \dots + y_k^2)^{\frac{1}{2}}$

It follows therefore from (12) that $\mu_\phi^{o}(q) = \mu_\phi^{o}(y_1)\,\mu_\phi^{o}(y_2)$, where $q = (y_1^2 + y_2^2)^{\frac{1}{2}}$. This can be written in the form

(314_1) $f(q_1 + q_2) = f(q_1)\,f(q_2)$; ($314_2$) $f(q) = \mu_\phi^{o}(q^{\frac{1}{2}})$,

where the q are arbitrary non-negative numbers. It follows that there cannot exist any $q = q_0$ satisfying $f(q_0) = 0$. This is clear from (314_2) and (12), if $q_0 = 0$. If, on the other hand, $q_0 > 0$, then successive application of (341_1) shows that $f(q_0) = 0$ implies $f(q_0/2^n) = 0$ for $n = 1, 2, \dots$; so that the zeros of f(q), and so, by (314_2), the zeros of $\mu_\phi(q)$, cluster at $q = 0$. But this contradicts (12), since, as proved in § 10, the function μ_ϕ^{o} is continuous. Thus f(q) is a continuous function which does not vanish for any q. In addition, f(q) is real-valued in view of (291) and (314_2). It follows therefore from (12) and (314_2) that f(q) is a continuous function which is positive for every q. Hence, (314_1) can be written in the form $g(q_1 + q_2) = g(q_1) + g(q_2)$, where $g(q) = \log f(q)$ is a real-valued continuous function. Consequently, there exists a real constant c such that $g(q) = cq$, i.e., $f(q) = \exp(cq)$.

This means, by (314_2), that $\mu_\phi^{o}(q) = \exp(cq^2)$. It follows therefore from (11) that c cannot be positive. Consequently, either c is 0 or the unit of length can be so chosen that c becomes $-\frac{1}{2}$; cf. (17).

In the first case, $\mu_\phi^{\,o}(q)$ is the constant 1. But then (15), (289) and (281) imply that the spectrum $S(\phi)$ consists of the single point $(0,\ldots,0)$. In the second case, $\mu_\phi^{\,o}(q)$ is identical with $\exp(-\frac{1}{2}q^2)$, which means, by (154), that μ_ϕ is the distribution θ having the density (205 bis). But then ϕ is the product distribution of k projections (287) each of which has the density (205 bis); so that ϕ is the k-dimensional distribution having the density

$$(315) \qquad \delta_\phi(r) = (2\pi)^{-\frac{1}{2}k} \exp\left(-\tfrac{1}{2}r^2\right), \text{ where } r^2 = (x_1^2 + \ldots + x_k^2).$$

This is Maxwell's "proof" for the validity of his law of the distribution of the velocities in a free ideal gas (the dimension number, k, is 3, the dimension number of the natural (x,y,z)-space). A more realistic approach is developed in Boltzmann's Vorlesungen. He considers a system of a large number of identical free molecules. Thus, if k is the degree of freedom, then, since the conservation of the energy means the conservation of the kinetic energy, the velocity components, say $x_1,\ldots,x_k$, are subject to the relation $x_1^2 + \ldots + x_k^2 = 1$, if the units are suitable chosen for every fixed k. Correspondingly, the energy surface would appear in the form $x_1^2 + \ldots + x_k^2 = k^2$, if the units were chosenthe same for k=1 as for anyk. The resulting discrepancy, introducing the reduction factor $k^{\frac{1}{2}}$ in the linear dimensions, will become fundamental in a moment.

Consider the equidistribution on the energy surface $x_1^2 + \ldots + x_k^2 = 1$. According to (297), the projection of this equidistribution has the density which is equal to $B_k(1 - x^2)^{\frac{1}{2}(k-3)}$ or to 0 according as x is or is not in the range $-1 < x < 1$. Hence it is clear that, if the radius of the sphere is changed from 1 to $k^{\frac{1}{2}}$, then the density of projection becomes $B_k k^{-\frac{1}{2}} (1 - x^2/k)^{\frac{1}{2}(k-3)}$ or 0 according as x is or is not in the range $-k^{\frac{1}{2}} < x < k^{\frac{1}{2}}$. Since $k^{\frac{1}{2}} \rightarrow \infty$ as $k \rightarrow \infty$, and since Stirling's (or Wallis') formula implies that the numbers B_k defined by (297 bis) satisfy the limit relation $B_k k^{-\frac{1}{2}} \rightarrow \pi^{-\frac{1}{2}}$, it follows that, as $k \rightarrow \infty$, the density belonging to the altered units tends to $\pi^{-\frac{1}{2}} \exp(-x^2)$ uniformly on every fixed bounded x-interval. But $\pi^{-\frac{1}{2}} \exp(-x^2)$ is a density which, except for a fixed change of the unit (in the proportion $1 : 2^{\frac{1}{2}}$), is identical with the density (205 bis).

Incidentally, it is seen from (304) that $A_k k^{-1} \rightarrow \pi^{-\frac{1}{2}}$, and that, in view of the structure of the function (304), the preceding approach to (205 bis) can therefore be applied without any change, if the k-dimensional sphere is replaced by the k-dimensional spherical space, considered in § 73. This latter approach to (205 bis), when translated into the language of the statisticians, means that the distribution of Student's ratio (303) for a large normal sample $(x_1, \ldots, x_k)$ is approximately normal, and that Student's result (304) is exact for a small normal sample as well.

REFERENCES

Monographs

[1] Aitken, A.C., Statistical Mathematics, University Mathematics Texts, no. 2, Edinburgh, 1929.

[2] Lévy, P., Calcul des probabilités. Paris, 1925.

[3] ______, Théorie de l'addition des variables aleatoires. Paris, 1937.

[4] Kolmogoroff, A., Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik und ihrer Grenzgebiete. Berlin, 1933, vol. 2, no. 3.

[5] Plummer, H.C., Probability and Frequency. London, 1940.

[6] Saks, S., Theory of the integral. Monografie Matematyczne, vol. 7. Warsaw, Lwow, 1937.

[7] Wintner, A., Lectures on asymptotic distributions and infinite convolutions. Ann Arbor, 1938.

The following notes are not of course intended to be complete. Their main purpose is to refer the reader to the original papers containing the principal general results on the one hand, and examples of theoretical interest on the other hand.

§26 - §41. The "Three Series Criterion" (§33) for convergence almost everywhere on the infinite product space is the fundamental theorem of A. Khintchine and A. Kolmogoroff, Recueil Soc. Math. de Moscou, vol. 32 (1925), pp. 668-677. Essential applications and interpretations of this general theorem were given by P. Lévy, Studia Math., vol. 3 (1931), pp. 119-155.

Lévy has, among other things, the fact that, in case of infinite product measures, convergence in measure implies (and is therefore equivalent to) convergence almost everywhere. In the language of infinite convolutions, this becomes the criterion given in §33; cf. B. Jessen and A. Wintner, Trans. Amer. Math. Soc., vol. 38 (1935), pp. 48-88. In case of a finite number of dimensions, the corresponding coalescence (viz., the identity of "independence" with "additive independence") does not take place; in this regard, cf. P. Hartman, E.R. van Kampen and A. Wintner, Amer. Jour. of Math., vol. 61 (1939), pp. 477-486. The theorem of §41 was proved by Jessen and Wintner (loc. cit.).

Historically, the whole development was initiated by Borel's formulation and proof of his "either 0 or 1" theorem (Rend. Palermo, vol. 29 (1909), pp. 247-271). Today it is easy, but at that time it was a true mathematical achievement, to think of the ordinary measure on the interval $0 \leqq x \leqq 1$ as a product measure in an infinite product space, the factors being the binary spaces corresponding to the dyadic expansion of x. It may be mentioned that not only the general theory of "random variables" but also Birkhoff's ergodic theorem are patterned after Borel's theorem, which indeed is a particular case (with metrical transitivity) of the ergodic theorem.

§42 - §48. Under the restriction to finite second moments, the determination of the Fourier transforms of semi-groups was carried out by A. Kolmogoroff, Rend. Accad. Lincei, ser. 6, vol. 15 (1932), pp. 805-808 and 866-869. The general theorem, that concerning arbitrary semi-groups, is due to P. Levy, Ann. di Pisa, ser. 2, vol. 3 (1934), pp. 1934; vol. 4 (1935), pp.

217-218. The above proof is based on the considerations used for finer purposes by P. Hartman and A. Wintner, Amer. Jour. of Math., vol. 62 (1940), pp. 757-779.

§49 - §52. Concerning the oldest criteria for the validity of Fourier's inversion formula, cf. the historical comments of A. Pringsheim, Jahresber. Math. Ver., vol. 16 (1907), pp. 2 - 16, according to which the first criterion applicable in case of conditional convergence is one given by A. Harnack (1885). The criterion of § 52 for continuity is nothing but the mathematical physicists' definition of a wave-length contained in the line spectrum; it may go back to Stokes (or to the school preceding him in optics). In connection with distribution functions as such, it was pointed out by P. Lévy [2], pp. 169-172. The Stieltjes form, (185), of Fourier's inversion and, first of all, the fundamental completeness theorem of § 9 are also due to him (loc. cit.; cf. Comptes Rendus, vol. 175 (1922), pp. 854-856 and Prace Mat.-Fiz., vol. 39 (1932), pp. 19-28).

The criteria (198), (200) for a high degree of smoothness were adapted to infinite convolutions by A. Wintner, Amer. Jour. of Math., vol. 55 (1933), pp. 309-331 and vol. 56 (1934), pp. 659-663; for further applications, cf. B. Jessen and A. Wintner, loc. cit. (and, in the latter connection, E. K. Haviland and A. Wintner, Duke Math. Journ., vol. 2 (1936), pp. 712-721). The existence of the symmetric "stable" distributions, those assigned by (202), was tacitly assumed by Cauchy for $0 < p < \infty$, and only Lévy observed the necessity, and has proved the sufficiency, of the restriction $p \leq 2$; for references, cf. A. Wintner, Amer. Jour. of Math., vol. 58 (1936), pp. 45-90, where a sharper result is proved. The expansion of § 55, where $0 < p \leq 1$, was given in Duke

Math. Jour., vol. 8 (1941), pp. 678-681.

§ 57 - § 68. In § 59, the mapping of the unit interval on the infinity product space of alternative choices is the idea which, in connection with Borel's paper, was pointed out above. This, the simplest, particular case of the general theorem of Khintchine and Kolmogoroff (1925) has been dealt with by R.E.A.C. Paley and A. Zygmund, Proc. Cambr. Phil. Soc., vol. 26 (1930), pp. 337-357. The infinite product (218) as such was introduced as a fairly general source of instructive examples (and counter-examples) by B. Jessen and A. Wintner, loc. cit., where the results of § 58 were also given. To § 60 - § 62, cf. F. Hausdorff, Math. Annalen, vol. 79 (1919), pp. 157-179; the application of Hausdorff measures to (217) was given by R. Kershner and A. Wintner, Amer. Jour. of Math., vol. 57 (1935), pp. 541-548. Those results of § 63 which concern an irrational r were proved by P. Erdős, ibid., vol. 61 (1939), pp. 974-976; cf. also vol. 62 (1940), pp. 180-186). As to § 64, cf. N. Wiener and A. Wintner, ibid., vol. 60 (1938), pp. 513-522 and Jour. of Math. and Phys. (M.I.T.), vol. 17 (1939), pp. 233-246. The estimates of § 66 were given [with the unnecessary restriction to the case (217)] by A. Khintchine, Math. Annalen, vol. 96 (1927), pp. 152-168.

§ 69 - § 75. The consideration of projections of distributions of radial symmetry goes back to Boltzmann's approach to the velocity distribution in ideal gases, subsequently considered by Borel. The above treatment is that given by P. Hartman and A. Wintner, Amer. Jour. of Math., vol. 62 (1940), pp. 757-779, where detailed references will be found to the literature of the subject.

Student's distribution is usually deduced as a property of normal distributions, rather than, as in §73, as a purely geometrical fact valid, for reasons of homogeneity, for any distribution of radial symmetry (whether the distribution be a product measure or not). For the corresponding metrical interpretation in terms of Riemann's spherical geometry, cf. A. Wintner, Astronomical Papers Dedicated to Elis Strömgren, Copenhagen, 1940, pp. 286-297.